1-1-1982

Transfer, study behavior and the pursuit of conceptual learning.

Philip N. Chase
University of Massachusetts Amherst

TRANSFER, STUDY BEHAVIOR AND THE PURSUIT OF CONCEPTUAL LEARNING

A Dissertation Presented

by

PHILIP N. CHASE

Submitted to the Graduate School of the
University of Massachusetts in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February, 1982

Department of Psychology

TRANSFER, STUDY BEHAVIOR AND THE

PURSUIT OF CONCEPTUAL LEARNING

A Dissertation Presented

by

PHILIP N. CHASE

Approved as to style and contend by:

Beth Sulzer-Azaroff, Chairperson of Committee

John Donohoe, Member

James M. Royer, Member

Richard Konicek, Member

Bonnie Strickland, Department Head
Department of Psychology

ACKNOWLEDGEMENT

The research reported in this manuscript arose from many sources and has been supported by many agencies and individuals. The first studies, those needed to develop the materials, were supported by the Psychology Department of the Fernald State School in Waltham, Massachusetts. In addition, my conversations with colleagues there lead to many revisions of strategy and purpose. I specifically would like to thank B. Barrett, C. Binder, L. Epstein, C. Hamad and R. Parry. Subsequent studies were supported by the Department of Psychology and the Mastery Learning Center of the University of Massachusetts/Amherst. The list of individuals from Amherst who contributed to this work is long, but I have especially profited from my interactions with E. Carlson, C. Fox, V. Hathaway, R. Konicek, D. Palmer, J.M. Royer, and W. Tyson. Preparation of the manuscript was supported by the Department of Psychology, St. Cloud State University. In particular, I would like to thank Lora Krueger for deciphering my crude drafts and typing the final copy. Finally, primary credit is given to the contributions of four individuals. I appreciate J. Donohoe's intellectual support at a number of critical junctures. I am indebted to K. Johnson for the collaborations and brainstorming that began these investigations. I am thankful for the motivation and clarifications supplied by B. Sulzer-Azaroff. And last, I am grateful for the emotional, intellectual, day-to-day support of K. Kenny. Without her contributions this manuscript would not have been completed.

ABSTRACT

Transfer, Study Behavior and the
Pursuit of Conceptual Learning

February, 1982

Philip N. Chase, B.A., University of Massachusetts
M.S. and Ph.D. University of Massachusetts

Directed by: Professor Beth Sulzer-Azaroff

These studies investigated the relations between different study programs and test performance. The study programs were comprised of combinations of prose passages and different types of study questions. Tests included four different types of questions. Both question types similar to those used in training (extension questions) and question types not trained (transfer questions) were included on the tests.

Experiment 1 investigated the differential effects of three kinds of questions. Programs included either definition, exemplify or example identification questions. Programs were prepared for three differen t concepts. Thus, subjects were exposed to three conditions composed of a different type of study question and concept. The results indicated that correct responding on example identification study questions was faster than on the other questions. However, performance was more accurate on definition and exemplify questions. On tests, correct responding was faster after example ident-

ification training than after the other training. Test performances after example identification and exemplify training were both more accurate than performances after definition training. In addition, extension performance was faster and more accurate than transfer performance. Significant concept differences were also found on both study and test performance.

Experiment 2 investigated the differential effects of three study programs. Programs included either example identification, both definition and example identification, or both definition and exemplify questions. Subjects were exposed to conditions composed of three programs and three concepts. The results indicated that correct responding was faster on example identification study programs than on the other programs. In addition, definition/example identification rates were higher than definition/exemplify rates. However, accuracy was higher on definition/exemplify programs than on either of the other programs. Analyses of other factors during study conditions revealed significant concept effects as well as a significant interaction between pretest or no pretest and the order of training. Analyses of test performance revealed: 1) faster and more accurate extension performance than transfer performance, 2) significant concept differences, 3) significant differences between exposure to pretest and no pretest, 4) significant

effects of order of training, and 5) significant interactions between pretest and concepts, and between pretest and study program. The results are discussed in terms of possible applications and further research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Recently, there has been a proliferation of research supporting the use of individualized instruction in secondary and post-secondary education. At least one basic system, The Personalized System of Instruction, has been developed for integrating the features of individualized pace, immediate feedback of performance, peer evaluation, increased student and instructor contact and criterion referenced evaluation (Keller, 1968; Johnson and Ruskin, 1977). Literally hundreds of studies and at least one meta-analysis review of these studies have demonstrated the success of the Personalized System of Instruction in higher education (Kulik, Kulik, and Cohen, 1979). However, as with any method of instruction, there are still aspects of individualized instruction that need improvement. One aspect that has received little attention is the design of study programs that facilitate learning complex, conceptual behavior.[1]

Individualized instruction relies on study programs to teach the content of a course or text. By reading a text, answering questions, and receiving feedback on the adequacy of the answers, the student learns the content covered by the questions. However, it is one thing to teach students to parrot the content of the text and quite another to instruct students to apply concepts appropriately in novel situations. Being able to recite the laws of motion would be insufficient for engineering students. They need to be able to use these laws to help analyze engineering problems and to design solutions. Certainly there are individualized courses that teach these kinds of higher level skills, but the rules for teaching such skills as problem solving, integration of desperate concepts and analyses of complex situations have not been delineated. Therefore, the general purpose of the studies reported here is to identify the rules for designing complex, conceptual study programs.

The following sections introduce the specific purpose and limits of this research. The experimenter poses an operant model of conceptual learning as an alternative to the two currently predominant models: the associationist and cognitive models. In order to clearly defend this perspective, all three models are defined, critical assumptions of each are specified and critical differences are explicated.

Purpose

This series of investigations was concerned with identifying those variables that influence the relations between individualized study materials and successful learning by college students. Specifically, these studies investigated how study programs comprised of different combinations of study <u>questions</u> might differentially effect those learning. The effects were tested by examining the subjects' responses to questions that were both similar to those included in the study program and those that were completely novel. The basic question was: which kind of study program facilitated test performance in response to the widest variety of questions.

This basic question may be conceived as a transfer of learning problem. Transfer of learning is defined as the performance on one task that influences performance on some subsequent task (Ellis, 1965). The subsequent task may be as similar to the first as repeating the same question. For example, a student is asked in a study program to define reinforcement. Later, the student is asked to define reinforcement on a quiz. If answering the study question makes it likely that the student answers the quiz question correctly, then transfer is facilitated. Such rote transfer is of little interest to most educators. However, transfer becomes more interesting as the initial and subsequent tasks become increasingly different from each other. For instance, a study question asks the student to define reinforcement. Later, the student is presented with the problem of explaining why a disturbed youth's outbursts in a classroom appear to increase in frequency. If the student stated that perhaps the teacher's attention for the youth's disturbances reinforced these outbursts, then transfer would have been demonstrated. In the first example, both questions or stimuli are identical and both answers or responses are identical. In the second example, both the stimuli and the responses are different. These two examples illustrate the extremes of a continuum of transfer across tasks for a particular concept. This dissertation examines this continuum.

Limits

The magnitude of this area of inquiry justified limiting these investigations to certain areas of transfer. Explicitly, these investigations were limited to verbal learning that is capable of being evaluated within a typical university setting. Verbal learning refers to behavior that is learned only through the mediation of other persons or the products of other people (Skinner, 1957). This behavior is typically divided into different topographical communication mediums: speaking, writing, gesturing. Since the experimenter decided to adopt procedures that could be used in a typical university classroom (i.e., one in which evaluation procedures are based on reading and writing), these studies are restricted to the medium of the writing. In addition, this investigation concentrated on transfer across types of behavior, but within a particular concept. Therefore, these studies emphasized the different responses that can be categorized by a single term. For instance, does performance on one type of verbal task such as defining a concept, transfer to performance on other kinds of tasks such as identifying novel examples of the concept? These studies were not designed to investigate transfer from a verbal medium to applications in the real world nor to investigate transfer from one concept to other concepts. It appeared that instances of such classes of transfer would be more difficult to program and evaluate than those that occur in the same medium and within the same concept. Since simpler classes of transfer problems had not yet been solved, these were addressed here. Later, when rules have been specified for the simpler cases, the more difficult problems of transfer should be studied.

Classifying Cognitive Tasks

Hierarchical schemes. One requisite to conducting these studies was to define the different kinds of verbal behavior that were to be taught and evaluated. Many investigators have defined, described, or otherwise classified cognitive or verbal skills. The most notable classification schemes are a taxonomy of cognitive objectives (Bloom, Engelhart, Furst, Hill and Kratwohl, 1956) and a hierarchy of learning types (Gagné, 1965). Both of these analyses have arranged cognitive behavior or tasks into hierarchial levels. The concept of level implies that tasks that are at the bottom of the hierarchy or taxonomy are prerequisites for tasks that are at the top of the hierarchy. For example, it is assumed that students must learn to verbalize definitions before they can sort positive and negative instances of a concept (Gagne, 1965). The concept of hierarchy implies that behavior at the top of the hierarchy subsumes comprehension level objectives (Bloom et al., 1956).

These classification schemes should be ideal for investigating transfer of learning, since they suggest that transfer occurs from higher level tasks to lower level tasks, but not from lower to higher level tasks. Since the lower level tasks are prerequisites, they would be subsumed by the higher. The instructional implications of these classification schemes are that if a student can complete higher level tasks, then it can be assumed that his behavior will transfer to lower ones. However, if a student fails a lower level task, it will be necessary to teach both lower and higher level tasks.

Unfortunately, neither of these instructional implications is supported by conclusive experimental evidence. Some studies have shown that it is necessary to acquire knowledge of a concept (memorizing a definition of a concept) before being able to sort successfully (Reed, 1946; Wolff, 1967). Other studies have shown that subjects can perform sorting and application tasks (applying the concept in new situations) without having to verbalize the definition of the concept (Furth, 1961; deLuna, 1972). In most cases it has been demonstrated that it is necessary to engage in the lower elements of different math hierarchies before one can engage in the higher elements of the math hierarchies (Gagné and Paradise, 1961; Gagné, Mayor, Garstens, and Paradise, 1962; Gagné, 1962; Gagné and staff, 1965). In these studies the exceptions to this rule were attributed to methodological flaws (White, 1973). For example, since pretesting was not used universally, it is possible that some subjects not trained in lower level behaviors could already perform lower level tasks (White, 1973). However, Kolb (1967, 1968) discussed the use of a similar methodology and the same hierarchy as Gagne' (1962) and found that many subjects succeeded at tests of higher elements while failing to learn relevant lower level tasks. The conclusions that can be drawn from such discrepant data are limited. As White (1973) indicated, both the studies that support the notion of hierarchies and those that refute this notion have had methodological problems. Therefore, the exact characteristics of a cognitive hierarchy remain unidentified at this time.

The fact that methodological problems have been cited as the cause of these inconclusive results is understandable. It may be true that conceptual hierarchies do exist, but that the method for testing the assumptions related to the hierarchy has not been implemented. This problem can be solved by utilizing alternative methodologies to test current hierarchies. However, there are other practical criticisms of these hierarchies that must be addressed first. Discussions with instructors, curriculum designers and educational researchers suggest that the levels of these hierarchies are complex to learn and use (Sullivan, 1969; Williams, 1977). In addition, it has been reported that content experts do not agree on the classification of objectives, or test items, within the categories of the hierarchies. (Sullivan, 1969; Williams, 1977). This suggests that even if one could demonstrate that a particular hierarchy does exist, the hierarchy would have to be re-written in order to be of any practical use.

Finally, both these and other verbal classification schemes (American Association for the advancement of Science, 1965; Gerlach and Sullivan, 1967; Williams, 1977) focus upon formal or topographical properties of behavior and neglect the functional history of the behavior. The classification of an instructional task should depend more upon the relation between the learner's behavior and previous instruction than on the particular words used in the task (Johnson and Chase, 1981). For example Markle and Tiemann (1970) discussed the following example:

> A long essay relating various trends (supposedly "synthesis") can represent rote learning, while a multiple-choice selection of the date of some event (supposedly "knowledge") can represent high-powered analytic thinking. (page 44).

The difference in complexity between these two tasks has more to do with whether the student has had contact with the answers/questions before or not. The long essay in the example may be a memorized verbal chain taken straight from an introductory text. The date may require relating three or four explorers, their discoveries and their historic era, yet none of this information has ever been related before in any text that the student could have read. The past behavior of the student and the teacher must be taken into consideration if instructional tasks are to be classified unambiguously.

In sum, the inconclusive results of research on cognitive hierarchies, the methodological problems, the problems related to agreement and the omission of the functional relations between student and teacher behavior all indicate that current cognitive classification schemes have proven inadequate to aid in defining the verbal behavior of interest. Therefore, it seemed necessary to construct a new classification system that minimizes these problems.

Concept Learning Research. The previous section argued that current schemes for classifying cognitive behavior are inadequate to investigate transfer of verbal learning problems. If that argument holds, perhaps other approaches might permit more adequate definitions of verbal behavior that could be used in experiments of transfer. Again, unfortunately the two predominant models of conceptual learning do not lend themselves to this purpose.

Basically, the two ways of categorizing previous research on conceptual learning are the environmental or associationist perspective and the cognitive perspective (Royer, 1979). The associationist perspective concentrates on defining the topographical similarity between stimulus events. Therefore, topographical analysis of the stimulus features are conducted and transfer or generalization of training is predicted on the basis of similar features or elements of the stimulus events. The cognitive perspective concentrates on the description of the internal mental processes that define memory. Transfer is predicted on the basis of the probability that certain information will be retrieved during a memory search process (Royer, 1979). In the following sections each of these perspectives are examined.

Associationist Perspectives. The vast literature covering associationist models of transfer has been criticized for its failure to yield a model of complex, conceptual behavior that is applicable to classroom instruction (Gagne and

Brown, 1961; Ausuhel, 1963; Clark, 1971; Royer, 1979; and Richards, 1979). These various criticisms can be divided into three basic contentions: 1) that there are fundamental differences between real-world learning and the learning that has been tested in laboratories influenced by associationist models 2) That the definitions of instructional tasks and of testable predictions are ambiguous and 3) that the exclusive concentration on the <u>antecedent stimulus</u> aspects of the environment can not account for all conceptual behavior. Ausubel (1963) argued that there are fundamental differences between real-word meaningful learning and the kinds of arbitrary stimulus-response relations studied by associationists. In addition, three reviews of basic research on conceptual behavior have suggested that perhaps the vast differences between laboratory research and classroom contexts for teaching conceptual behavior contribute to the paucity of classroom applications of laboratory findings. Although none of the reviews specifically focused on associationist literature, most of the studies reviewed could be categorized by that label. Clark (1971) logically analyzed both research and classroom contexts and found five major differences. First, most of the research used separate conjunctive concepts where as classroom instruction requires relational conjunctive concepts (e.g. teaching the dichotomy chair, not chair as opposed to the relation between chair, chaise lounge and sofa). Second, research subjects have usually been required to attain already familiar concepts; several of these simultaneously. In contrast, the classroom student is required to learn new, unfamiliar and successive concepts. Third, the concept instances presented in research tasks have been almost exclusively concrete, have more than one feature, the features were discrete (e.g., black or white), the number of features was finite and the features were absolute (i.e., they were either there or not there and didnot change over time). On the other hand, the concept instances presented in classrooms are often abstract, have any range of features, the features are continuous (light blue-blue-dark blue), the number of features can be infinite and the features and dimensions may change over time. Fourth, the research strategy for teaching concepts has most often been the discovery method; the subject discovered the critical features, values and so on. However, in the classroom, the most frequently used strategy is expository; the student is told the critical features to be learned. Finally, concept attainment in concept research was most often evaluated by a sorting task (either verbal or physical manipulation) while in the classroom students are asked to define or use the concept (Clark, 1971, pp. 254-255).

Gagné and Brown (1961) also speculated that the difference between laboratory concepts and real-world concepts prevented clear-cut application of experimentally derived principles to the classroom. The concepts acquired in the course of an experiment are not further used as in the solution of a problem, but are simply measured as being established. This point is similar to the final point made by Clark discussed above; that the tasks used to evaluate subjects performance in basic research are different from the tasks used to evaluate students performance in classrooms.

Finally, Rickards (1979) analyzed the results of recent research conducted on inserting questions within prose passages. He claimed that all of these studies employed somewhat artificial procedures since subjects were not allowed to check the text passage when answering the inserted questions. Therefore, it would be inappropriate to generalize to the natural environment in which students are reading textual materials and can always refer back to previous pages to find the passage or concept in question (Rickards, 1979, pp. 193).

A more serious criticism of associationist models is that these perspectives lend themself to vague predictions, vague procedures and thus, vague understanding of conceptual behavior (Chomsky, 1959, 1971).

For example, Chomsky (1971) stated:

> When we look for specific predictions (about the environmental effects on behavior) we find virtually nothing ...
> What does it mean to say that a sentence of English that I have never heard nor produced before belongs to my repertoire...
> Skinnerians[2] appeal to "similarity" or "generalization," but always without characterizing precisely the ways the new sentence is "similar" to familiar examples or "generalized" from them.
>
> -Chomsky, 1971
> pp. 20

Chomsky (1971) later stated that the reason such similarities can't be specified is because the similarity is not in the environment, and therefore can only be expressed by referring to or postulating internal states of the organism (e.g. a grammar). This point of view will be discussed in a later section, but in any case it is clear that defining the similarity between events is not entirely understood.

The third criticism is closely related to the preceding conclusion (Chomsky, 1959, 1971): that associationist models can not account for some examples of conceptual behavior. Royer (1979) presented an instance of a real-world problem that may not have been encountered in the classroom, yet the tasks learned in the classroom could be and often are applied to the solution of the problem. Suppose a child learns to compute the area of a rectangle in a classroom. After instruction the child is faced with the problem of determining the amount of carpet needed to cover a living room floor. The relevant question is: does the instruction and the problem share enough environmental features to predict that the child will solve the problem? If the child does solve the problem and has never been presented with examples of measuring carpets before, how can an associationist model account for the behavior? It is not clear that in all cases of transfer, especially those involving verbal rules, that sufficiently similar features are available in both the learning and the testing environment to predict transfer.

Royer's (1979) example is just one of many similar cases in which generalization or transfer occurs when the similarity between one antecedent event and a subsequent event are not simply the perceived or observed similarities. Other examples include Chomsky's (1971) English sentence example quoted earlier, Rickard and Denner's (1978) -ed rule misapplication by young children and even Russell's (1927) discussion of Watsonian language theory (i.e., two different sentences may have nothing physically in common, yet relate the same fact). All of these examples suggest that simply specifying the relation between antecedent stimulus events is not sufficient for predicting whether the response will occur or not.

Thus to summarize the criticisms of research conducted from an associationist perspective: First, it was established that there are vast differences between the kinds of learning that have been tested in laboratories and that of the real world. Second, both the model and its related research have been shown to yield vague predictions and ambiguity in specifying instructional tasks and strategies. Third, associationist models appear to inadequately account for some forms of conceptual learning. Therefore, it is concluded that neither a classification system of conceptual behavior nor a research strategy should be based exclusively on this particular model.

Cognitive Perspectives. Because associationist models have failed to solve the problems of predicting and controlling conceptual learning, many researchers have turned to cognitive models. In these analyses, the problem of transfer becomes one of retieval of relevant knowledge that has been stored in internal structures. However, these models have also been criticized on the basis of their underlying assumptions.

Cognitive models hold at least five basic assumptions about conceptual learning and performance. First, that internal mental activity exists. Second, that this mental activity is structurally connected to the perceptual systems. Third, that mental activity is somehow different from observable behavior. Fourth, that this mental activity can be logically inferred on the basis of environmental input and behavioral output. Finally, that this mental activity controls observable behavior.

Many critics of cognitive models would agree to some of these assumptions if they were accompanied by certain qualifications (cf. Skinner, 1969). For instance, if it is agreed that mental activity is the same as neurological activity then the first and second assumptions are reasonable. However, it is not clear that this neurological activity adheres to laws that are different from other behavior. In fact, one could postulate a neurological model that is as functional as a behavioral model (cf. Trehub, 1977).

However, stating such a hypothesis or any other hypothesis about the general structure of the neurological activity related to conceptual behavior is not the critical problem of cognitive models. Rather, the critical problem lies in the methods used for testing such hypotheses. Cognitive models assume that such claims can be tested by specifying the logical relations between input, output and the hypothesized internal state, and by conducting experimental studies to examine these relations. This assumption is difficult to support.

The first problem with this assumption has been suggested by Anderson (1978). Anderson (1978) described a debate that has centered on one question of cognitive science: what is the nature of representation in memory? Does representation consist of images or propositions? Anderson (1978) argues that there is no way to distinguish between kinds of representation on the basis of behavioral data. Given the same input, different theories of representation predict the same behavioral outcome simply by postulating different process variables. Other criteria such as parsimony (e.g. fewest features needed to account for the most phenomena), plausibility (e.g., model does not contradict common experience), and efficiency (e.g., the latency between input and output takes the least amount of time), may be of some assistance, but without the basic knowledge of the physical dimensions of neurological activity these criteria are all subjective. Parsimony and efficiency are relative, and plausibility is simply another name for common sense. The conclusion is that one cannot determine whether representation is imaginal or propositional. Since all investigation of human learning uses behavior as its basic datum and since representation, a critical component of cognitive theory, can not be determined by looking at behavior, cognitive models of conceptual behavior appear to be seriously limited.

This criticism of cognitive research is a specific example of a general critique of a mentalistic theory. Even though one can account for the input and output variables of a system, there are an infinite number of routes that might mediate these events. Consider a geometric analogy. Two points in space may be connected by a line, but in order to determine exactly whether the points are connected by a straight line or some kind of curve, more than two points are needed. There are an infinite number of lines that connect two points. Likewise there are an infinite number of possible ways that a stimulus and a behavior can be mediated. Since other criteria, such as parsimony and efficiency, can not be applied to unknown qualities, such as the speed of electro-chemical reactions to verbal stimuli, the methodology used in cognitive research is probably less than optimally efficient.

Ironically, the computer lends itself, by analogy, to another criticism of cognitive assumptions. Palmer (1980) suggested that if one were to teach a college sophomore to use a computer it would be best not to concentrate on the internal mediating processes of the computer. Rather, one should concentrate on the operations the sophomore would perform and the feedback obtained from the computer. Furthermore, if one were to teach this sophomore the internal mechanisms of the computer, one would not suggest that the sophomore look at a series of inputs and outputs to try to deduce these mechanisms (as we have seen above this might take forever). Instead, it might be best to begin by either taking the computer apart piece by piece and putting it back together or by designing a schematic program that would demonstrate some of the critical features of the computer. Thus, both the function and the structure of the computer can be accounted for by direct, observational techniques.

If this analogy seems to be familiar, it is because it takes the same form as the age old story of the first empiricist. A group of monks were sitting around a long table, discussing how they can determine the number of teeth in a horse's jaw. After days of such deliberation, a young upstart raised his hand and asked, "Why don't we find a horse and count his teeth?" Of course, the young monk was banished from the room and from all further investigations of natural phenomena. (Johnston and Pennypacker, 1980). Certainly, determining the neurological activity that is part of the verbal process is a more difficult phenomenon to investigate than horse's teeth. However, the level of difficulty does not necessarily imply that a different methodology should be used. If neurological activity is to be described, then direct investigation of neural components of learning is needed.

Many authors have argued against these kinds of criticisms of mentalism and inferential reasoning. Specifically, Lachman, Lachman & Butterfield (1979) have argued that the adequacy of a cognitive theory can be determined by <u>convergent validation</u>. Convergent validation is a process by which the adequacy of a model is either supported or refuted by the results of many different kinds of investigations. If these results or conclusions converge, the model is validated. However, Palmer (1980) stated three conditions under which the use of convergent validation would lead to erroneous conclusions:

> 1. If one's model is circular or is so general that it can account for all conceivable results of an experiment.
>
> 2. If one's model includes elements of a simpler and sufficient model.
>
> 3. If one's model simply describes a higher order process that tends to be conditioned in all subjects owing to universal controlling variables.
>
> Palmer, 1980
> pp. 2

Under all these conditions, the model might be adequately described without revealing anything about the universal controlling variables. Thus, convergent validation is one criterion for the consistency of a model, but it is insufficient to eliminate the problems described earlier.

The final point in this critique of cognitive approaches is the assumption that mental events control behavior. If control implies that neurological structures limit an organism's interactions with certain environmental arrangements, then certainly no argument can be tendered. In these cases, the preceding argument for neurological study as opposed to cognitive study is germane. However, it is not clear that this definition of control is universally accepted by cognitive theorists. Some appear to claim that cognitive events produce behavior Chomsky (1965). In these cases, the problem of infinite regression must be argued (Ryle, 1949). For example, when it is said that someone has "acted intelligently" and then it is claimed that this intelligence was controlled by corresponding mental activity, one might then inquire if this mental activity was done intelligently. If so, then one would have to postulate another activity to accompany the first. Then, one could ask the same question of this and all other mental events, indefinitely. If our task is to discover how instruction influences conceptual learning, it would seem that even if the mental activity that accompanies conceptual learning was determined, one would still need to look at the environment for the events that control learning.

In sum, it has been contended that some of the basic assumptions of cognitive perspectives are open to question. It appears that inferences made on the basis of two observable events could well lead to an infinite number of indistinguishable, yet potentially different hypothesized structures. In turn, little information on how to arrange instruction would be derived from these postulated structures. In addition, it has been argued that even if the mental activity that accompanies conceptual behavior could be determined through cognitive methodologies, the problem of infinite regression would hinder any attempts to identify controlling relations. It would still be necessary to investigate the environment in order to derive practical solutions to problems of learning.

Thus, it has been argued that neither the associationist nor the cognitive models of conceptual learning are sufficient. Therefore, an alternative strategy for defining conceptual behavior and for investigating the variables by which it is controlled must be undertaken.

An operant model of conceptual learning. The operant model is an environmental model that is directly related to the previously discussed associationist model. However, it includes more aspects of the environment than the associationist models described. An operant account of learning does not attribute sole causality of learning to the relations between antecedent stimuli. In addition to these important relations, operant accounts look at the relations that exist between antecedents, behaviors and consequences. The unit of interest is a three-term functional unit: stimuli that precede a response, the response itself and the effect that the response has on the environment.

Skinner (1957) described a model of verbal behavior that was based on an operant account of learning. Skinner (1957) categorized verbal behavior on the basis of functional distinctions. For example, one class of verbal behavior, intraverbal, was defined as any verbal behavior, spoken, written or gestural, that is controlled by a spoken, written or gestural verbal stimulus that does not have point-to-point or one-to-one correspondence with the behavior. A tact, on the other hand, is verbal behavior that is controlled by a physical or non-verbal stimulus. Transcriptive behavior is a written response that has point-to-point correspondence with a written stimulus. The intraverbal, the tact and the transcriptive define relations between stimuli and responses that are functionally different. For the sake of isolating the different kinds of control exerted over verbal behavior, these classes are presented as discrete units. However, in the natural environment, most complex verbal interactions are composed of combinations of these and other classes of behavior.

How does this Skinnerian or radical behavioral approach differ from those models that have failed to produce adequate direction? First, consider the associationist models of conceptual learning. Associationism was criticized for three reasons:

1) the differences between research conducted from this perspective and real-world learning, 2) the vague

predictions and ambiguous instructional implications of this research, and 3) the inadequacy of this model to account for certain classes of learning. Conceptual learning research conducted from an operant perspective differs from these other environmental perspectives in a number of ways.

First, operant research of human learning has typically dealt with real-world learning. The research on both programmed instruction and behavioral instruction (e.g., PSI) was motivated by training and teaching needs that existed in the armed services and in large universities. These behavioral procedures were used to teach real concepts within classrooms and other traditional learning environments (cf. Holland and Skinner, 1961; Keller, 1968). Similarly, the investigations described in this dissertation were motivated by practical considerations and a laboratory procedure was created that closely approximates a real world learning activity. First, actual concepts were placed within the context in which they usually appear, within prose passages. Second, during study trials, subjects were allowed access to text passages and to their answers to previous questions. However, during test trials they were not. Third, answer keys were provided to the subjects after they had attempted an answer to a study question. Of course, some differences existed (e.g. the continuous presence of an experimenter). However, every attempt was made to create a learning environment that approximated the real world.

Second, most specific applications of the operant model do produce clear and precise predictions and controlled technologies. Although Gagné and Brown (1961) and others have suggested that programmed instruction as defined by Skinner (1958) and Galanter (1959) does not necessarily facilitate learning, Markle (1967) and Holland & Kemp (1965) did indicate specific steps that should be taken in order to create successful programs. Similarly, Keller's description of PSI (Keller, 1968; Keller and Sherman, 1978) has been sufficiently specific to be replicated and the predictions about PSI's effectiveness have been reproduced a number of times (Kulik, Kulik and Cohen, 1979). It is simply premature to comment on the adequacy of research that specifically deals with conceptual behavior from an operant perspective. The critical research in this area has yet to be completed. In fact, the investigations reported here are designed for the purpose of leading toward such predictions and procedures. All efforts have been made to be as specific and unambiguous as possible. The adoption of a functional, operational classification system described in simple, concrete terms is one step towards promoting specificity.

Third, operant models of learning do not concentrate solely on the associations between antecedent stimuli.

As stated earlier, one reason that environmental theories have been deemed inadequate is that the critics have concentrated on the problems of predicitng antecedent-antecedent relations. This is understandable. Most research on concept formation, verbal learning and transfer of learning conducted from an environmental perspective has emphasized the similarity of two or more antecedent stimuli. However, antecedent-antecedent associations are one of many kinds of environmental relations. In order to account for these various relations operant theorists consider a number of other variables. All of the relations that exist between antecedents, behaviors and consequences are important. To concentrate on one of these sets of variables to the exclusion of the other two and to the exclusion of historical or temporal relations such as the arrangements of scheduling of these variables is insufficient.

As stated earlier, one of the most common discussions of the inadequacy of environmental models centers on the use of general grammatical rules by humans. One of the best examples of this phenomenon is the misapplication of the -ed past tense rule that is often observed in young children. Interesting enough, Skinner (1957) used this particular example to demonstrate the effect of the reinforcing community on the verbal behavior of the individual (Keenan and Grant, in press). A child hears many sentences in which the speaker discusses the past, and the verb or action in these sentences often ends with the sound "-ed". Although the child probably never hears the word "teached", when he begins to verbalize past events he may very well say "teached." How can this be? If one concentrates on determining the similarities between stimuli that may have been antecedent to this behavior, one is hard pressed to come up with an adequate explanation. Perhaps some other similar words like "reached" or "impeached" have some control, but since the child also says "ised" and "bended" etc. it seems unlikely that verbs similar to teach are responsible. However, if one applies even the simplest notion of a reinforcement schedule and includes antecedents, behavior and consequences in an analysis, the existence of this behavior is not such a mystery.

For instance, it must be remembered that none of this behavior occurs in a vacuum. The notion of reinforcement schedules is important because it attempts to account for patterns of behavior. The child has been involved in many verbal interchanges before this misapplication takes place. His correct use of past tense verbs with "-ed" endings has probably been reinforced repeatedly. In addition, he has heard many more verbs used with an "-ed" ending than any other form of the past tense. It is not too far fetched to assume that he has even heard some adults misapply the -ed rule and still be reinforced or not punished. One also assumes that any arrangement which is repeated or made more salient than other arrangements probably changed the child's behavior in some way. Thus, given the similarity of the antecedent-behavior-consequence relations and the frequency or density of reinforced "-ed" arrangements, one would predict that a child would use it with many verbs. The three term relation that has occurred frequently is similar to the misapplication situation even though the specific antecedent may not be.

It is hoped that the above example adequately describes now an operant analysis augments the associationist perspective. Although similar in its focus on environmental variables, operant analyses are different from associationist perspectives because the major criticisms of associationism do not apply. Thus, an operant approach may well prove to be a viable model on which to base the pursuit of an environmental model of conceptual learning.

An operant model of verbal learning also differs from cognitive models of learning in several ways. Briefly, cognitive models were criticized on two counts: 1) the problem of inferences based on input and output and 2) the problem of infinite regression when internal events are said to control external events. Certainly behavioral models do not encounter either of these problems. External events are considered the ultimate source of control and inferences are made only about the probability of similar stimulus-response-stimulus relations occurring in the future.

In order to describe the advantages of an operant model over other previously described models for classifying conceptual instruction, further developments are necessary. Certainly, research based on this perspective of conceptual learning has been slow to materialize. However, several texts in addition to Skinner (1957), Winokur, 1976; Peterson, 1980), the development of programmed instruction, and behavioral systems of instruction (e.g., PSI), and an organization that has nurtured the notion of a functional account of language (Association for Behavior Analysis) have combined to help generate a model of conceptual instruction. Specifically, this model starts with a general educational goal: the need to teach students to engage in the behavior of professionals or advanced students in a discipline (Markle and Tiemann, 1970; Bostow, 1976). In other words, in order to isolate the complex, conceptual behavior that is important to teach students we need to look at the behavior of the content experts. Content experts engage in various kinds of discourse with respect to the materials within their discipline. Experts can state the facts and figures of their discipline. They can relate seemingly obscure similarities between concepts. They can identify real world instances of the concepts developed in their disciplines. They can provide students with intriguing examples of these concepts. Finally, they can ask questions and determine methods for answering these questions, or when faced with a problem can determine ways of solving the problem; in short, experts can problem-solve. The sum of these various classes of verbal behavior constitutes the complex verbal repertoire needed in advanced educational and professional environments. The particular components of this sum constitute what is necessary to teach students in secondary and post secondary instructional settings.

Explicit in these educational goals is the classification system for verbal behavior that was developed by Skinner (1957). Skinner's classification system permitted educators to define the specific kinds of verbal behavior that might be of interest to teach. In order to use the verbal classification system for instructional purposes; Johnson and Chase (1981) designed a typology of verbal instructional tasks that is based on Skinner's (1957) analysis. The definitions of the tasks in the typology describe functional relations between stimuli and behavior and the same labels are used that Skinner (1957) coined (i.e., echoics, transcriptives, intraverbals, tacts). However, the typology also includes types of questions that appear to be discussed by others (Anderson, 1972; Andre and Sola, 1976; Andre and Biddle, 1978; Frase, 1968; Watts and Anderson, 1971). Appendix B defines and exemplifies five of the tasks from the typology. Table 1 shows the extent of the typology.

- - - - - - - - - - - - - - - - - -

See TABLE 1, Page 33-34

- - - - - - - - - - - - - - - - - -

Having reviewed a classification system of conceptual instructional tasks based on an operant model it is possible to compare the typology to the problems described with other classification systems. The major criticisms were: 1) that the hierarchies were not supported by experimental evidence, 2) that they did not consider either the function of the tasks for the learner or the history of the learner with respect to the tasks, and 3) that there was low agreement between experts on the classification of objectives and questions.

Criticisms 1 and 2 are immediately countered by the definitions specified in Appendix B. The typology does not make any claims about levels of difficulty or hierarchial arrangements. In addition, each category specifically discusses the functional relation between teacher behavior and student behavior. The distinction between <u>elementary</u> and <u>conceptual</u> behavior also clarifies the problem of the learners history with respct to the tasks.

The third criticism was empirically tested. Chase (1980) examined the typology to determine whether others

TABLE 1

TYPOLOGY OF VERBAL INSTRUCTIONAL TASKS

Elementary[1] and Conceptual[2] Tasks	Examples
Echoic	Correctly repeat the following lines from Shakespear's Hamlet. Be sure to copy my intonation closely.
Textual	Correctly pronounce the following medical terms:
Transcriptive	
Copying from text	Correctly copy the following Chinese letters.
Taking dictation	Correctly spell the following names for laboratory equipment as I say them.
Intraverbal	
Definition	Define reinforcement.
Example Identification	Say which of the following written scenarios is an example of positive reinforcement:
Exemplify	Give an example of reinforcements.
Tact	
Example Description	Describe the technical properties of the plant specimens onthe laboratory test table.
Example Identification	Say whether each of the following video-tapes scenarios illustrates assertive or aggressive behavior:
Example Component Analysis	Identify at least three distinctive features of each of the wines in the goblets in front of you.

TABLE 1 (continued)

Combinations	Any two or more of the above tasks. Includes tasks requiring mands.

----

1 Require fixed verbal behavior

2 Require flexible, extended verbal behavior

could reliably classify tasks according to the definitions given in Appendix B. After developing a program to teach the five task types, graduate and advanced undergraduate students of psychology learned to categorize test items according to the typology. The mean time for completion of the study program was 21 minutes. Then, each subject was given a classification test compared to twenty novel items. Ten of the items were taken from the experimenter's own course materials (Introductory and Educational Psychology). The other ten were taken from commercially available materials. Subjects were asked to identify what type of task each items illustrated. The mean performance on the classification test was 88% agreement with the experimenter. Since both the study time and the agreement scores were better than those found with other classification systems (Williams, 1977) the investigator was convinced that a reliable typology had been developed.

In conclusion, this section has argued that there are many practical, methodological and logical problems with previous attempts to investigate conceptual behavior. These problems have interferred with developing a set of classroom prescriptions for teaching conceptual behavior. However, this section has also contended that an operant model could well succeed where the others have not. The operant model of Skinner has certainly contributed to the development of a classification system of verbal instructional tasks. Once this classification system was developed, it was possible to develop materials to investigate the questions about transfer of learning. The answers to these questions may lead to rules for writing individualized study materials. The following chapters describe two studies that were designed to accomplish this purpose.

CHAPTER II

EXPERIMENT I

Purpose

The first experimental test of transfer of learning examined the differential effects of study programs that consisted of single task types. Three classes of tasks were compared; example identification, exemplify and definition. The copy task was not compared because a number of authors reported that similar task types (i.e., verbatim or memorization tasks) do not facilitate learning conceptual behavior (Ellis, 1965; Andre, 1979; Johnson and Stratton, 1966; Watts and Anderson, 1971; Miller and Weaver, 1976; Keenan and Grant, 1979). Copy tasks were used, however, as the first question in each study program. They functioned as an observing response "to make sure that subjects had read the passage." The combination task was not compared because of its complexity and composition. Since the combination task by definition combines two or more task types into one, it was reasoned that the separate effect of each should be investigated prior to investigating the parameters of the combination task. Combination tasks were used, however, as test items.

The specific questions asked by this initial investigation were: does study performance or acquisition vary across the three types of tasks and does test performance on a variety of tasks vary according to the type of task learned?

Methods

Subjects. Twelve undergraduates recruited from a special Introductory Psychology course for transfer students at the University of Massachusetts served as subjects. All subjects were sophomores and juniors majoring in Psychology and had mastered introductory level concepts in both basic learning principles and experimental methodologies before the experiment took place.

Personnel. A graduate student in Educational Psychology coordinated the study and two undergraduate psychology majors served as research assistants. The research assistants conducted the experimental sessions, corrected the tests, and transcribed tapes. The experimenter trained the assistants and checked the reliability with which the experimental procedures were carried out. Research assistants were trained in both experimental procedures and on the concepts used in the study programs. Experimental training consisted of: 1) a detailed written description of the correct procedures for each session, 2) modeling by the experimenter, 3) role-playing and 4) feedback from the experimenter. Both assistants met the criterion of no more than one procedural mistake during a role-playing session before participating in the study. Training on content consisted of: 1) studying the prose passages for each concept, 2) answering all tasks, 3) feedback on performance, 4) reanswering all tasks that were incorrect, and 5) terminal feedback. Both assistants met the criterion of 95% correct performance on these tasks before participating in the study.

Setting. The study was conducted in two similar sound-insulated carrels. Each carrel was equipped with a desk, two chairs and shelves for experimental materials. A one-way mirror connected the two carrels.

Materials and apparatus. The experimental materials included three prose passages each of which defined an esoteric psychological concepts (abulia, constructional approach and tau effect). For each concept a copy task, a set of examples and nonexamples, a series of definitional questions, a series of exemplify questions and two problem solving or combination questions were written. (See Appendix C for examples). All of these materials had been previously tested for difficulty. Chase, (1980), developed a methodology for creating sets of instructional materials that could be compared for level of difficulty. Using a combination of techniques developed by Markle (1967), and Merrill and Tennyson (1974) for testing programmed materials, Chase (1980) was able to develop and revise the programs and test materials used in these experiments. Individualized feedback was obtained on such components as step-size, length of programs, humor, grammar and other editorial concerns. These data were used to rewrite the prose passages and questions. The level of difficulty was determined by the proportion of subjects who answered questions correctly. A question was considered difficult if between 15-20% of the subjects answred it correctly. A question was considered easy if 90-100% of the subjects answered it correctly. Questions considered too difficult were those that were answered correctly by less than 15% of the subjects. These were eliminated. Questions that were easy were used only in the study programs. Difficult questions were used as terminal study and test questions. In this way, the influence of individual task difficulty across conditions could be minimized.

In addition to determining the difficulty of individual tasks, these procedures also determined whether the three concepts used in the study, abulia, constructional approach and tau effect, were of equal difficulty. A post-hoc comparison of concept effects revealed no statistical differences between abulia and tau effect. However, the constructional approach was found to be significantly more difficult than both abulia and tau effect, $F\ (2, 15) = 4.42$, $p < .05$. Although the program for constructional approach was revised many times these differences were maintained throughout the studies. As discussed in chapter 3, these differences persisted as a source of variability and produced compelling data in their own right.

In addition to these materials, there was a study behavior questionnaire and accompanying prose passage (Johnston, O'Neill, Walters & Rasheed, 1975; Chase, 1980) (Appendix D), a pretest (Appendix E), and a comment and scoring sheet (Appendix F). Each carrel was also equipped with procedural outlines for each session. Assistants used these outlines to conduct the experimental sessions. All materials were typed on 8½ x 11 paper and photocopied.

Cassette tape recorders were used to record all interactions between research assistants and subjects. At the beginning of each session the assistants recorded their names, the subjects' names, the date the session number and the concept being studied. These tapes were used to check the reliability with which the procedures were implemented.

An electric timer was also provided for each carrel. The duration that subjects spent on each task was recorded by the research assistants. The timer was also used during the reliability probes to rescore the duration of each task.

<u>Procedures</u>. The study was conducted with each individual subject during four one-hour sessions. The first session was devoted to assessing the subjects' entering behavior

with respect to study skills, and the concepts <u>abulia</u>, <u>constructional approach</u>, and <u>tau effect</u>, by means of the study behavior questionnaire and pretest.

The study behavior questionnaire asked the subjects to estimate the amount of time that they spent engaged in various study behaviors while studying the prose passage. For instance, while they were reading a 900 word passage they were asked to keep track of how much time they spent on the first reading, on the second reading, on underlining and on making notes. As indicated in Table 2, subjects were asked to use their typical study techniques. They were told before starting that they would be tested on the content of the passage.

The pretest consisted of sixteen questions. There were four questions related to the content of the prose passage used for the study behavior questionnaire and four types of questions on each of the three concepts, <u>abulia</u>, <u>constructional approach</u> and <u>tau effect</u>. The first question type asked the subject to state a definition of the concept in their own words. The second type requested that examples of the concept be identified. The third asked for an original example of the concept and a fourth type posed a problem that required defining, identifying and exemplifying the concept in the context of a scenario. Thus, these questions were parallel to the questions that were later asked on the tests for each concept. Table 2 outlines the procedures that assistants followed during the first session.

- - - - - - - - - - - - - - -

See TABLE 2 Page 44-45

- - - - - - - - - - - - - - -

The second through fourth sessions were similar to one another. Each session was different only with respect to the type of study condition manipulated. These differences are specified in Table 3. During each session the general format was as follows:

- - - - - - - - - - - - - - -

See TABLE 3 Page 46

- - - - - - - - - - - - - - -

First, subjects read a prose passage that defined a concept. The assistant placed the passage on the desk in front of the subjects and turned on the timer. When subjects finished the passage, the timer was turned off and the duration was recorded. Second, the assistant presented the copy task. Subjects were told to fill in the blanks of the copy task word-for-word from the passage. The assistant left the passage on the desk for subjects to copy. Again, as soon as the task was presented to the subjects, the assistant turned on the timer. When the subjects had completed the task, the timer was turned off and the time

TABLE 2

EXPERIMENTAL PROCEDURES FOR THE PRETEST AND STUDY BEHAVIOR QUESTIONNAIRE SESSION

PLEASE FOLLOW THESE DIRECTIONS CAREFULLY

I. Before the subject arrives:
   A. Bring the scoring sheet, the 900 word passage, scrap paper, the study behavior questionnaire and the pretest to the table.
   B. Set the timer to zero and assure that it is working.
   C. Test the tape recorder. Write the tape number, side number and counter number on the scoring sheet.
   D. Write your name, the subject's name, date, concept and session type on the scoring sheet.

II. After the subject arrives:
   A. Say hello, have a brief informal discussion about anything.
   B. Describe the study in general, including a description of the tape recorder and timing procedure.
   C. Turn on the tape recorder.
   D. Describe this session.
   E. Show the S.B.Q. to the subject; point out each of the questions; explain their purpose; to find out how students typically study for a test. Answer any questions.
   F. Show subject the 900 word passage.
   G. Show subject how to time each of the study skills while studying the passage.
   H. Answer questions. Tell subject to knock on window (one-way mirror) when ready to be tested on the passage.
   I. Set timer and leave carrel.
   J. When subject knocks, turn timer off and record the duration of study time.
   K. Check each item and question subject to assure that all relevant items were answered.
   L. Show the subject the pretest; describe each type of question.
   M. Ask subject to read each question carefully and answer it as completely as possible. Tell subject that if the answer is not known, to write "D.K." for Don't Know next to the question.
   N. Answer any questions. Ask subject to knock on window when done. Give the test.
   O. Set timer and leave the carrel.
   P. When subject knocks, turn off timer and record duration.

TABLE 2 (continued)

Q. Check to see if each question has been answered.
R. If not, ask subject to complete the questions. Turn timer back on.
S. If questions are all answered, thank subject and make an appointment for the next session.

TABLE 3

DESCRIPTION OF THREE STUDY CONDITIONS REGARDLESS OF CONCEPT

Define Condition

A.	Prose passage defining the concept	A.	Prose passage defining the concept	A.	Prose passage defining the concept
B.	Copy task as observing response	B.	Copy task as observing response	B.	Copy task as observing response
C.	Define task on feature 1 with prose passage	C.	Exemplify task on feature 1 with prose passage	C.	Example of concept (with passage)
D.	Define task on feature 1 without prose passage	D.	Exemplify task on feature 1 without passage	D.	Nonexample of concept that varies feature 1 (with passage)
E.	Define task on feature 2 with passage	E.	Exemplify task on feature 2 with prose passage	E.	Nonexample of concept that varies feature 2 (with passage)
F.	Define task on feature 2 without passage	F.	Exemplify task on feature 2 without passage	F.	Nonexample of concept that varies feature 3 (with passage)
Terminal Define task (complete definition) without passage		Terminal Exemplify task (complete original example) without passage		Sequence of examples and non-examples of concept without passage	

Tasks on the same feature are parallel, not identical.

Subject must respond correctly to every question before progressing to next question.

Exactly half the questions were answered with the passage available, for the referral by the subject. The other half were answered without the passage.

recorded. Next, the assistant immediately corrected the copy task. If there were any mistakes, the subjects were asked to correct them.

Upon completion of the copy task, the assistant presented the series of tasks for the specific program that was assigned for that session. Each task was presented separately and was timed, recorded, and corrected. The timing procedure was identical to that described above. The subjects also received detailed, specific feedback for each answer based on a prepared answer key. The answer was read and mistakes if any, were indicated. If the answer was correct, the subjects were told why. These procedures were followed until the study sequence was completed. The assistant then asked if there were any questions. If not, the test for that particular concept was given. The test consisted of a series of examples and nonexamples, two terminal definition questions, two exemplify questions and two combination or problem solving questions. Table 4 details the construction of each transfer test for each concept.

- - - - - - - - - - - - - - - - -

See TABLE 4 Page 48

- - - - - - - - - - - - - - - - -

The test administration followed the same pattern as the study sequence: each task was presented separately and the duration that the subjects spent answering the

TABLE 4

DESCRIPTION OF THE TRANSFER TESTS FOR EACH CONCEPT REGARDLESS OF STUDY CONDITION

constructional approach	abulia	tau effect
A. 11 example discrimination tasks; two nonexamples for each varied feature, three examples	A. 9 example discrimination tasks; two non-examples for each feature, three examples	A. 15 example discrimination tasks; two nonexamples for each varied feature five examples
B. 8 define tasks; each feature asked for twice	B. 6 define tasks; each feature asked for twice	B. 8 define tasks; each feature asked for twice
C. 8 exemplify tasks; each feature asked for twice	C. 6 exemplify tasks; each feature asked for twice	C. 8 exemplify tasks; each feature asked for twice
D. 2 combination tasks; each broken into four parts	D. 2 combination tasks; each broken into four parts	D. 2 combination tasks; each broken into four parts

question was recorded. However, during the test sequence subjects were not given any feedback. After they finished one task, the next was immediately given. This sequence continued until all transfer tasks had been completed.

At the end of each session the assistant asked the subjects to sign up for the next session. If it was the last session, the assistant provided written feedback for the study and asked the subjects to check with the experimenter for specific feedback. At this point the subjects' involvement in the experiment was completed.

Experimental Design. An intrasubject, repeated measures design was used. Each subject was trained with each of the study programs, studying each of the three concepts: (Constructional approach, tau effect, and abulia. All subejcts were randomly assigned to one of three counter-balanced conditions. A Latin square counterbalanced assignment of subjects to a sequence of study programs was used to control for any treatment order effects. Concept order effects were not manipulated because of the difficulty of analyzing data with both order of treatments and order of concepts varied. Table 5 illustrates this design.

- - - - - - - - - - - - - -

SEE TABLE 5 PAGE 50

- - - - - - - - - - - - - -

An intrasubject design was used because it was possible to look at the effects of independent variables on the per-

TABLE 5

SEQUENCE OF EXPERIMENTAL CONDITIONS FOR EACH SUBJECT

Subject	Session 1 Abulia	Session 2 Tau effect	Section 3 Con. App.
1	Exemplify	Definition	Example ID
5	Exemplify	Definition	Example ID
9	Exemplify	Definition	Example ID
12*	Exemplify	Definition	Example ID
2	Definition	Example ID	Exemplify
6	Definition	Example ID	Exemplify
7	Definition	Example ID	Exemplify
10*	Definition	Example ID	Exemplify
3	Example ID	Exemplify	Definition
4	Example ID	Exemplify	Definition
8	Example ID	Exemplify	Definition
11*	Example ID	Exemplify	Definition

* Indicates subjects eliminated from analyses because of substantial experimenter error in implementing the procedures.

formances of each individual, not just the central tendency of a group of individuals. The increased data and control afforded by the intrasubject analyzed allowed the investigator to probe the nature of any functional relations that might have existed between individual subjects' performance and the independent variables. The critical comparison of treatment effects was made within each subject. Therefore, the subjects served as their own controls. Technically, each additional subject, in this case eight, was used for purposes of direct and systematic replication. (Sidman, 1960). Each time subjects received the exact same order of study programs, this constituted a direct replication. Direct replications demonstrate whether the experimental treatment produced a robust effect as a function of the treatment (Sidman, 1960). Through direct replication generality across subjects is determined. Each of the three different sequences of treatments constituted systematic replications. Systematic replications demonstrate whether an experimental treatment exerted powerful control despite planned variations of procedure (Sidman, 1960). Through systematic replication, generality regardless of sequence or order effects could be determined. Intrasubject designs with direct and systematic replications allow for generality to be established for individuals, not just groups of individuals. Since the purpose of this study was to compare the effects of different study programs on the text performance of individuals, individual generality is important. If between subject variability is great then the data suggest the need to look at other variables.

Reliability and interscorer agreement. Two tactics were used to determine if experimental procedures were reliably implemented. First, approximately 10% of the experimental sessions were scored for procedural agreement or accuracy. The agreement index for the pretest session was determined by observing whether assistants followed the procedures outlined in Table 2. Items IIC, D., E., G., H., K., L., M., N., and R., from Table 2 were considered essential to the experiment. The experimenter listened to the tape of the session and when the assistant engaged in one of these behaviors a "+" was scored. The number of pluses was divided by the total number of items and multiplied by 100 to obtain percent agreement. The mean agreement for eight rescored pretest sessions (four per assistant) was .87.5 with a range from 70-100 percent agreement and a median of 88%. Agreement indices for the other sessions were more difficult to ascertain as assistants often did not discuss what was occurring in the session. Therefore, 10% of the tapes were replayed by the experimenter simply to determine whether any extraneous information was given to the subjects. In three cases, additional information critical to the experiment was discussed. For example one assistant paraphrased the definition of constructional approach for a subject. In all three cases these and other substantial departures from the experimental protocals required eliminating the subjects' data from analyses of the results.

Second, approximately 45% of all the experimental sessions were rescored to calculate indices of interscorer agreement for answers to study questions. Low agreement on these indices meant that subjects had received incorrect feedback. Therefore, their scores on the transfer tests after such feedback would not be indicative of the study condition they had received. The interscorer agreement indices were calculated for each session by dividing the number of agreements by the total number of items scored and multiplying by 100. The mean agreement was 92.25% with a range of 66 - 100% agreement and a median of 94%. Two of the three cases eliminated from further analyses were eliminated because of low agreement indices (a 66% agreement and a 77% agreement). Therefore, for data used in subsequent analyses, the mean interscorer agreement index was 94.5%.

Interscorer agreement indices were calculated for each dependent measure. Performance on pretest and transfer test accuracy was calculated by scoring each as "correct" or "incorrect" with the number of correct responses divided by the total number of illustrations. The definition tasks were scored on the basis of 20 points. Each answer was divided into 10 parts; each part worth one point. Therefore, if the subject's responses were entirely correct on both define tasks, the number of points given, 20, was divided by the total number of possible points, 20. Both the exemplify and the combination-task scores were calculated in the same way. The answer keys used for these scores are in Appendix G. The agreement indices were calculated by rescoring 25% of all the data. Agreement was defined as the number of agreements divided by the total number of items that had been rescored, multiplied by 100. All agreements on example-identification scores were 100%. Mean agreement on definition scores was 93.37% with a median of 100%, mean agreement on combination scores was 88.3% with a median of 93% and mean agreement on exemplify scores was 89.6% with a median of 100%. The range for all of these agreement indices was 60 - 100%.

In addition to these measures of interscorer-agreement, an index of interrecorder agreement was calculated for the recording of durations. A second observer observed 8 of the 42 experimental sessions through the one-way mirror connecting the two carrels. The second observer activated the timer when the subject was engaged in each task and recorded the time it took the subject to finish the task. These times were compared to those recorded by the research assistant. Durations were considered to be in agreement if they were within $\pm$ 2 seconds of each other. The agreement indices were calculated by dividing the number of agreements by the total number of durations scored for each session. The mean agreement was 86.75% with a range of 68-96% and a median of 84%.

Results

This section presents the effects of three independent variables: study programs, concepts and order of presenting the study programs. The analyses were separated into the general dependent variables of study performance and test performance. Study performance analyses were further separated into the dependent measures of rate of correct performance, percent correct performance, number of errors and duration. Test performance analyses were further separated into the dependent measures of rate of correct performance, percent correct performance and duration. Rate of correct test performance and percent correct test measures were analyzed in terms of total test performance, performance on extension items (those test items that correspond to the class of tasks that was trained) and performance on transfer items (those classes of tasks that had not been trained).

Test performance was separated into these components for three reason. First, total test performance was the most meaningful measure because it included all four classes of tasks and because it was most analogous to a classroom evaluation. Second, extension performance was isolated to determine whether the specific learning acquired the sequence was maintained on the text. Third, transfer performance was isolated to answer the question: how well does training on one class of tasks transfer to other classes of tasks. Duration were reported only for the total test. In addition, both study and test performance were analyzed in terms of both intrasubject and group comparisons. Other possible relations are presented in the discussion section.

Pretest and study behavior questionnaire data were stored for future analyses. When a sufficient number of studies have been completed to permit an adequate examination of these data, these analyses will be conducted.

Study performance. Figures 1-3 present the correct responses per minute (slope), duration (length of responding) and accuracy data for all nine subjects. Individual graphs display subjects' data according to the combination of study program and concept which they received. The black circles represent correct responses and the open circles represent incorrect responses. The time in minutes indicates the time subjects spent answering the questions (duration). It does not include the time required to present the questions nor the time required to give the subjects feedback on their answers. The Y axis is a scale of the cumulative number of correct responses given.

First, Figures 1-3 demonstrate that constructional approach was more difficult for subjects to learn than sequence was maintained on the test. Third, transfer performance was isolated to answer the question: how well does training on one class of tasks transfer to other classes of tasks. Durations were reported only for the total test. In addition, both study and test performance were analysed in terms of both intrasubject and group comparisons. Other possible relations are presented in the discussion section.

Pretest and study behavior questionnaire data were stored for future analyses. When a sufficient number of studies have been completed to permit an adequate examination of these data, these analyses will be conducted.

Study performance. Figures 1-3 present the correct responses per minute (slope), duration (length of responding) and accuracy data for all nine subjects. Individual graphs display subjects' data according to the combination of study program and concept which they received. The black circles represent correct responses and the open circles represent incorrect responses. The time in minutes indicates the time subjects spent answering the questions (duration). It does not include the time required to present the questions nor the time required to give the subjects feedback on their answers. The Y axis is a scale of the cumulative number of correct responses given.

First, Figures 1-3 demonstrate that construction approach was more difficult for subjects to learn than either of the other concepts. Eight subjects' rates of correct responding were lower on constructional approach than either abulia or tau effect (Subjects 1-4; 6-9). The sole exception (Subject 5) was the only subject to score above 90% correct on all three study programs and the only subject who claimed to use all three kinds of questions when "typically" studying for a test (Study Behavior Questionnaire data). Therefore, this subjects entering skills and performance were not typical. A planned comparison between constructional approach and both abulia and tau effect did not yield a significant difference, $F (1, 12) = 1.75$, $p < .05$. Other relations, though, showed constructional approach to be more difficult than the other concepts. The lowest performances on definition and exemplify tasks occurred with constructional approach and two of the slowest example identification performances occurred with constructional approach. Error analyses revealed that the most errors on all study programs occurred with the constructional approach. Finally, analyses of duration demonstrated that the constructional approach programs took longer than the abulia programs for all subjects and longer than tau effect programs for eight subjects (Subjects 1-4; 6-9). A planned comparison among concept durations yielded a significant difference between constructional approach and the other concepts, $F (1, 12) = 27.79$, $p < .01$.

- - - - - - - - - - - - - - - -

SEE FIGURES 1-3, PAGES 59-66

- - - - - - - - - - - - - - - -

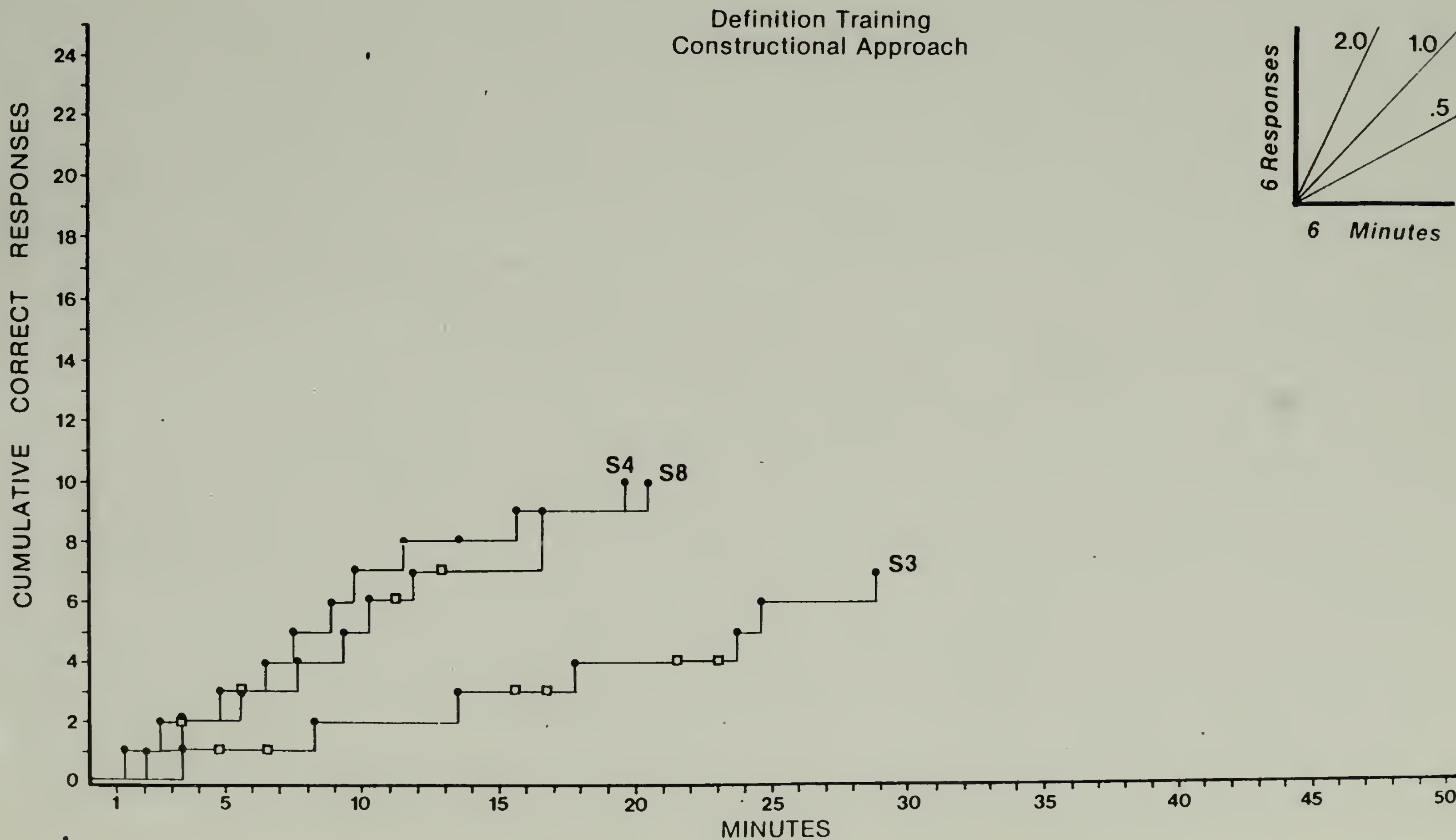
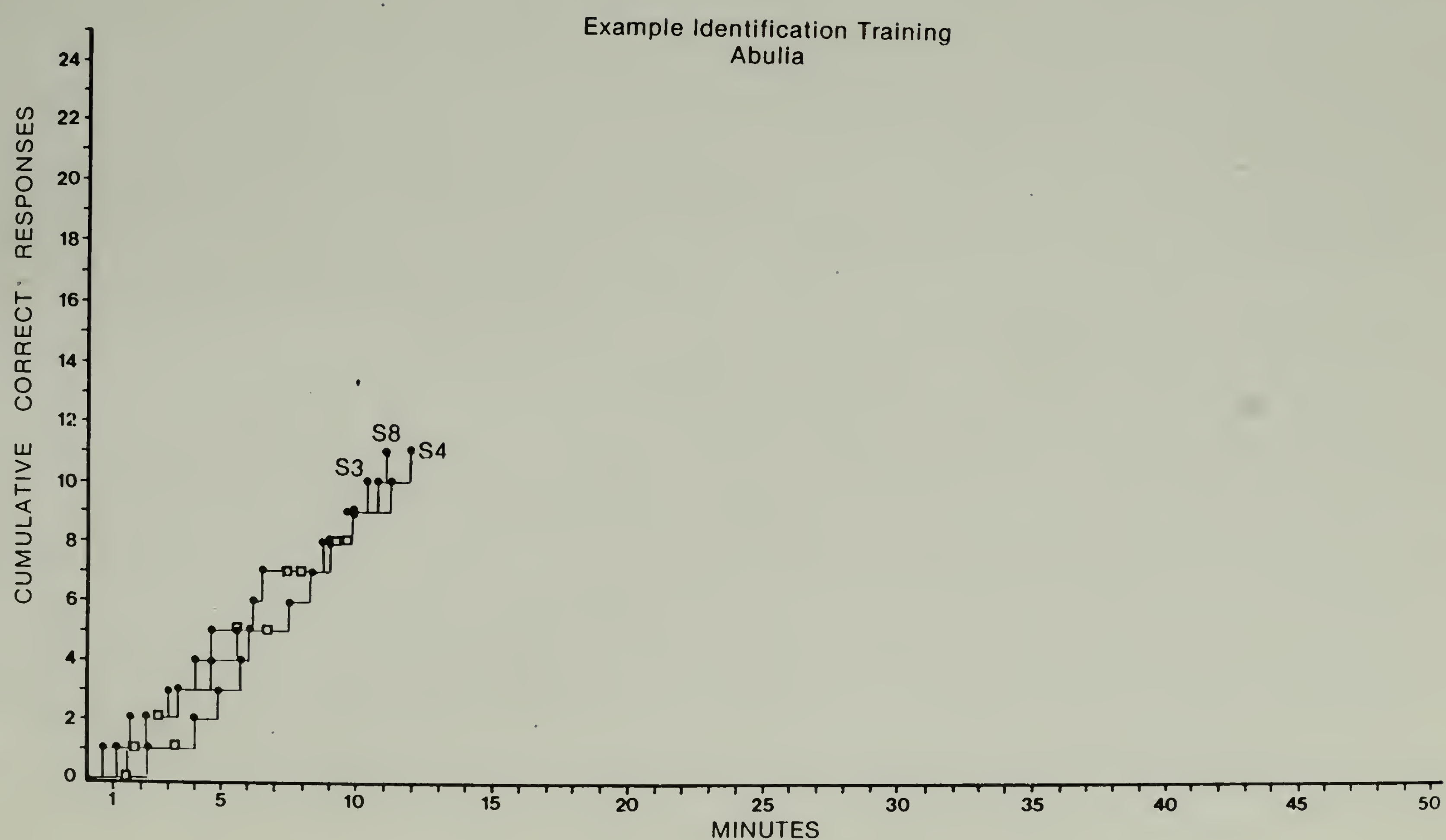
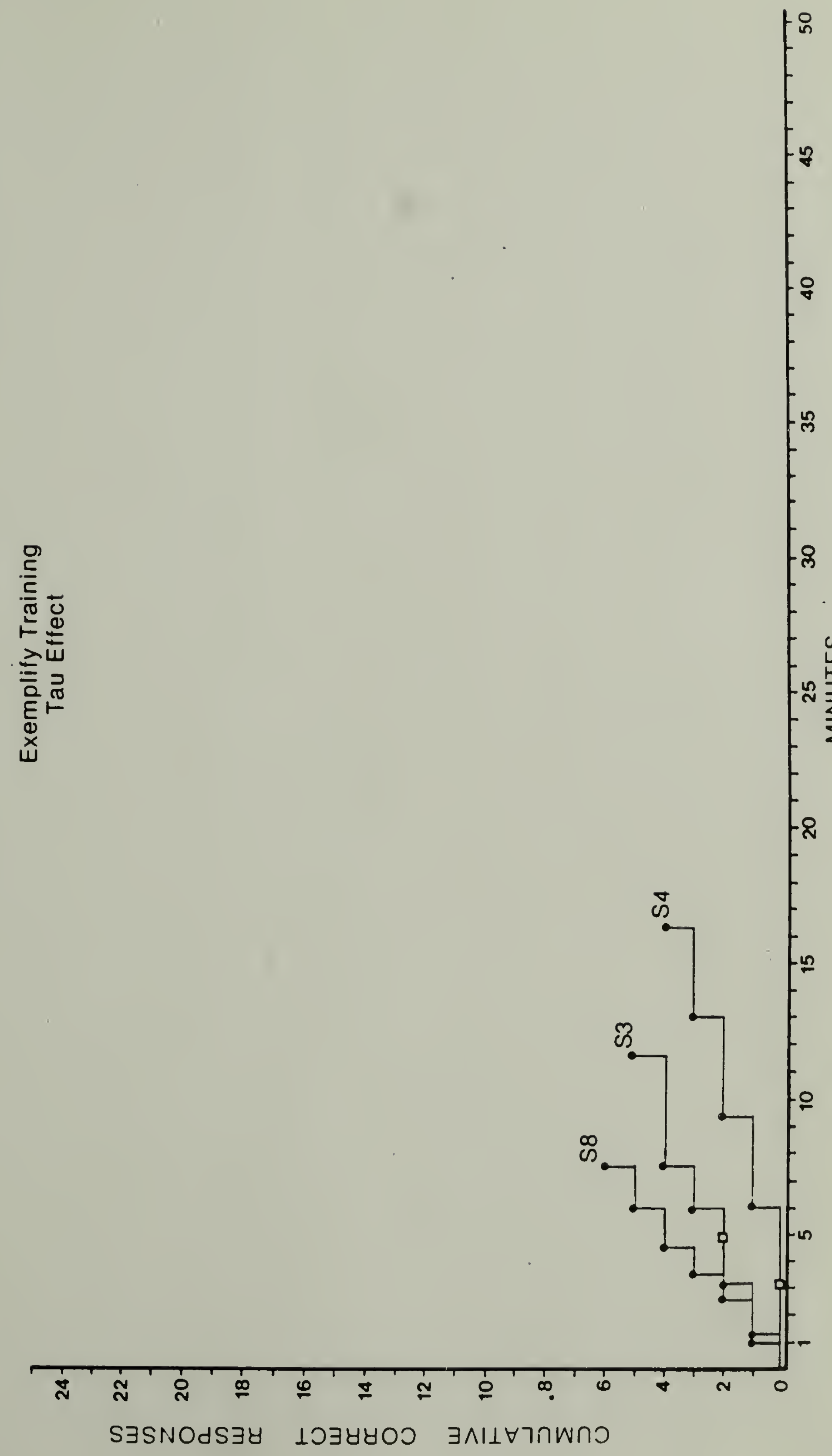


Figure 1. Cumulative frequency of correct study performance.


Example Identification Training
Abulia
CUMULATIVE CORRECT RESPONSES
MINUTES
S3
S8
S4
0
2
4
6
8
10
12
14
16
18
20
22
24
1
5
10
15
20
25
30
35
40
45
50



Exemplify Training
Tau Effect
CUMULATIVE CORRECT RESPONSES
0
2
4
6
.8
10
12
14
16
18
20
22
24
MINUTES
1
5
10
15
20
25
30
35
40
45
50
S8
S3
S4

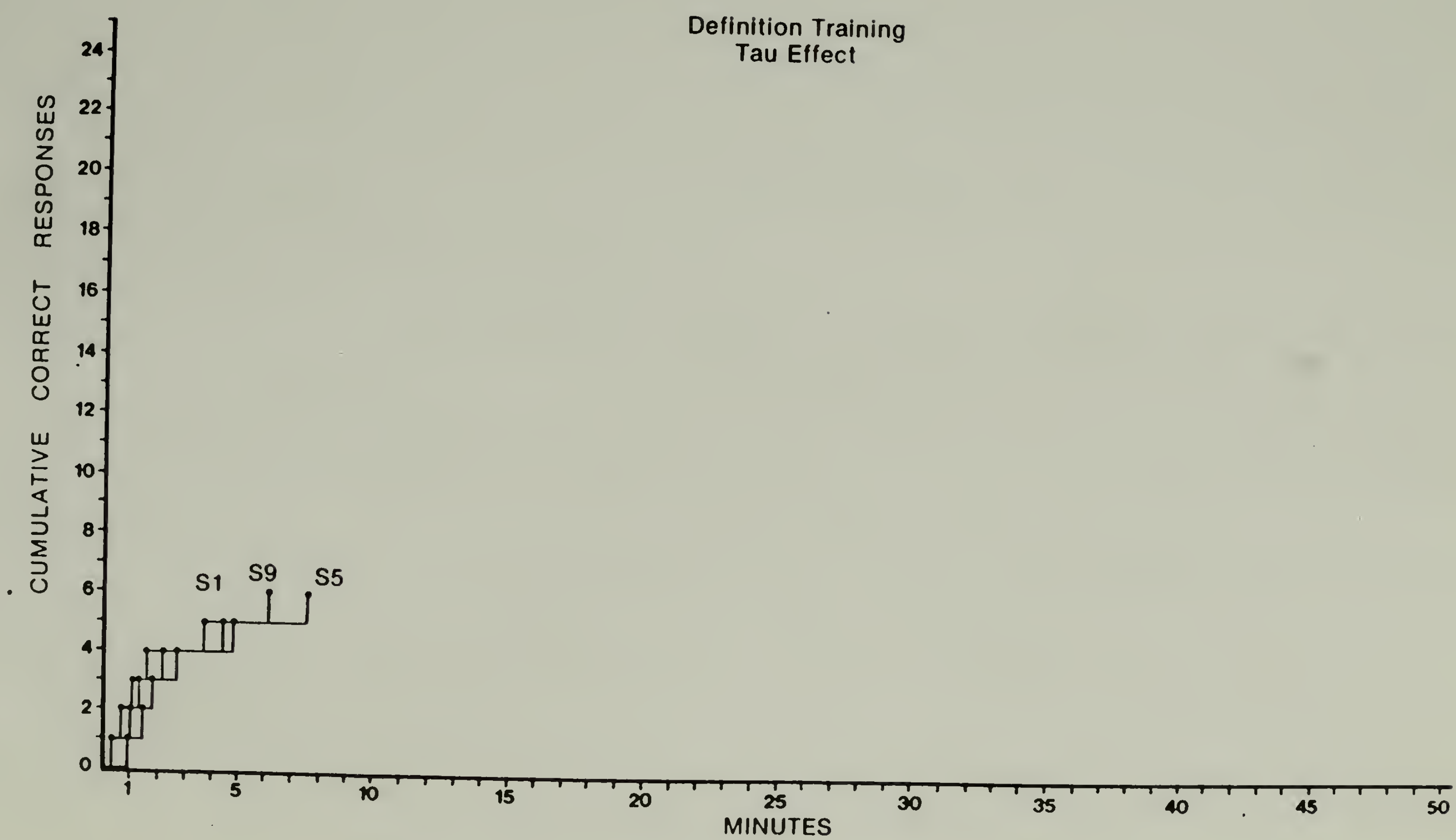
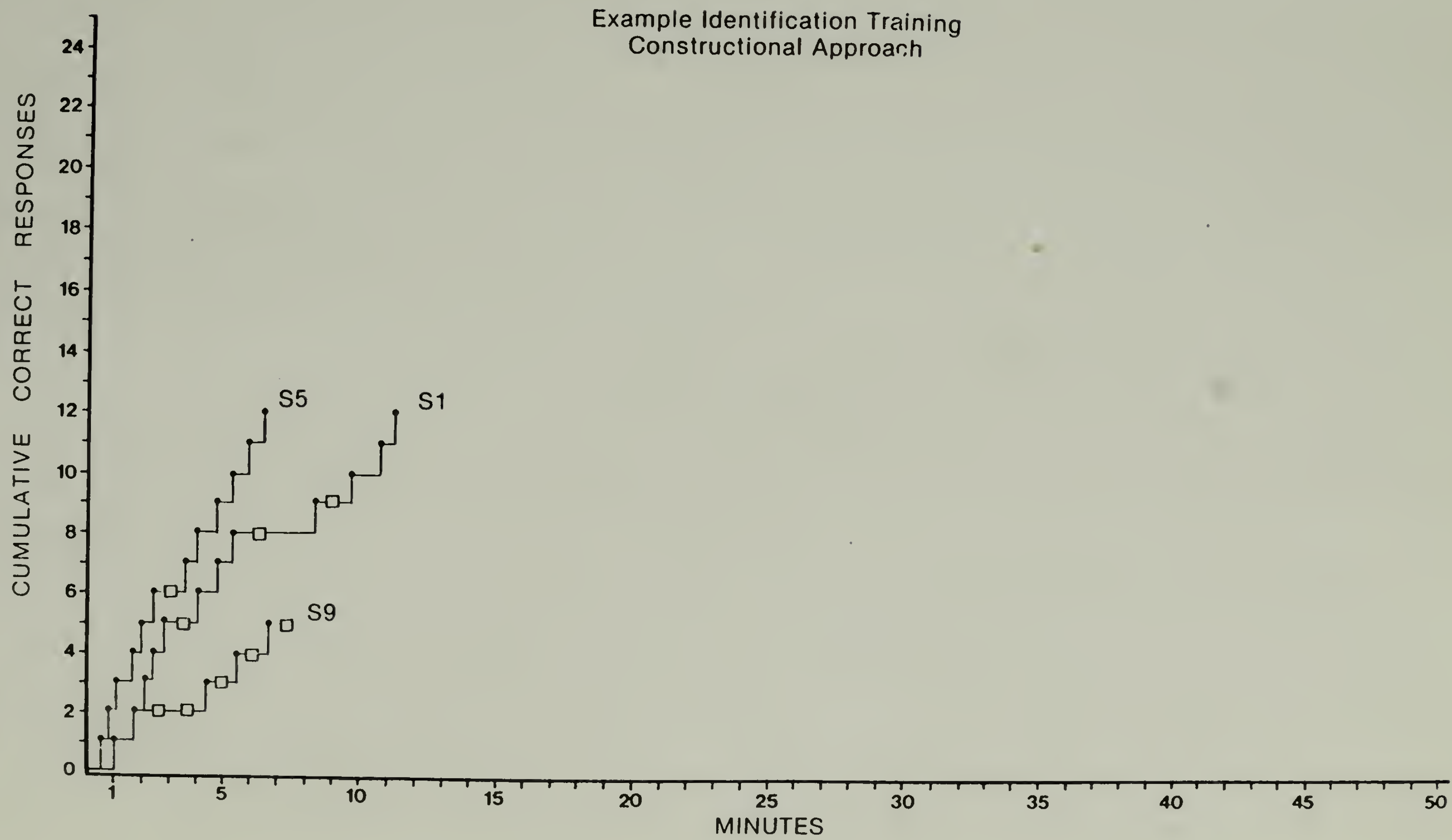
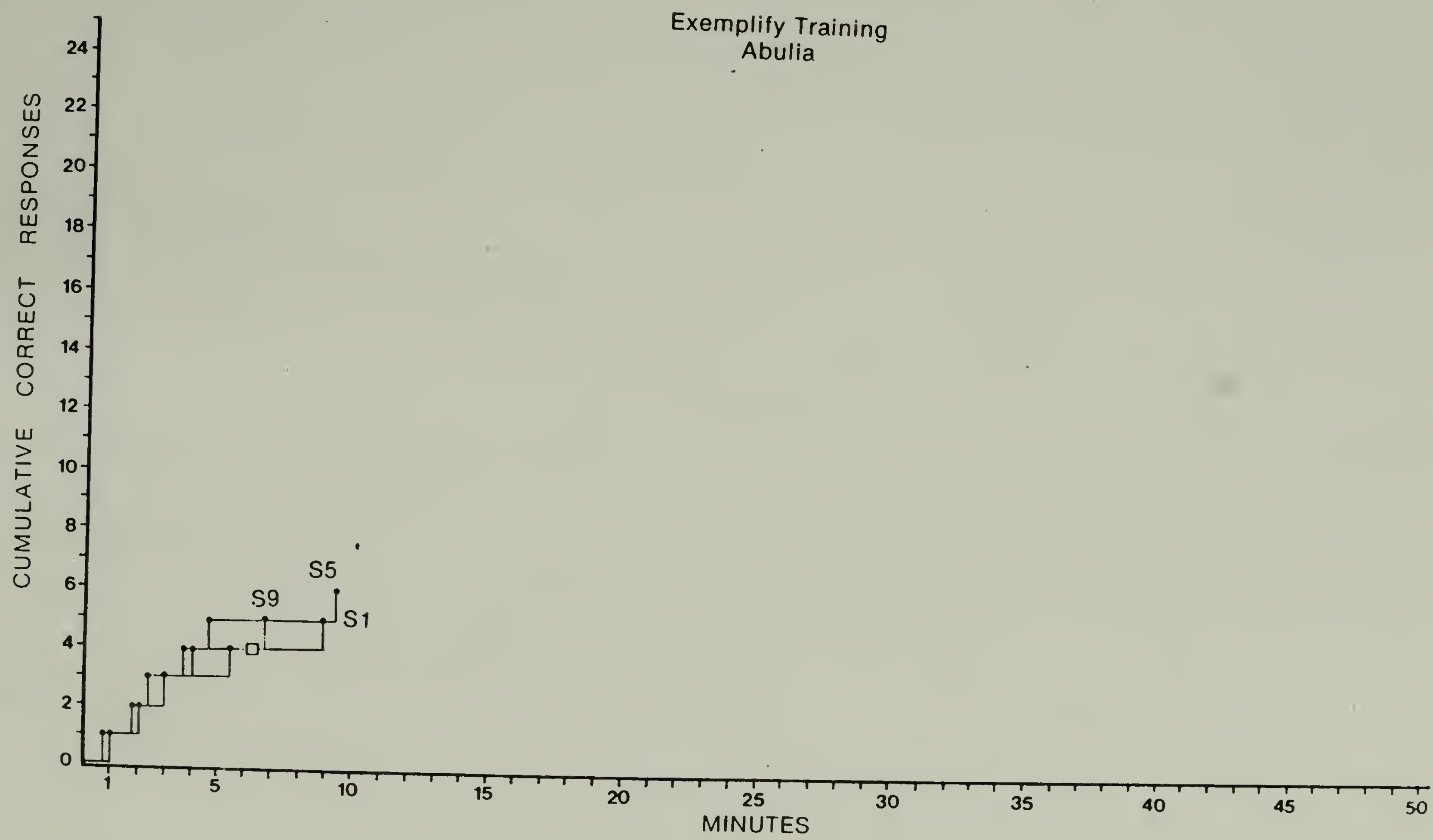


Figure 2. Cumulative frequency of correct study performance.


Example Identification Training
Constructional Approach
CUMULATIVE CORRECT RESPONSES
0
2
4
6
8
10
12
14
16
18
20
22
24
S5
S1
S9
1
5
10
15
20
25
30
35
40
45
50
MINUTES


Exemplify Training
Abulia
CUMULATIVE CORRECT RESPONSES
MINUTES
S5
S9
S1
0
2
4
6
8
10
12
14
16
18
20
22
24
1
5
10
15
20
25
30
35
40
45
50

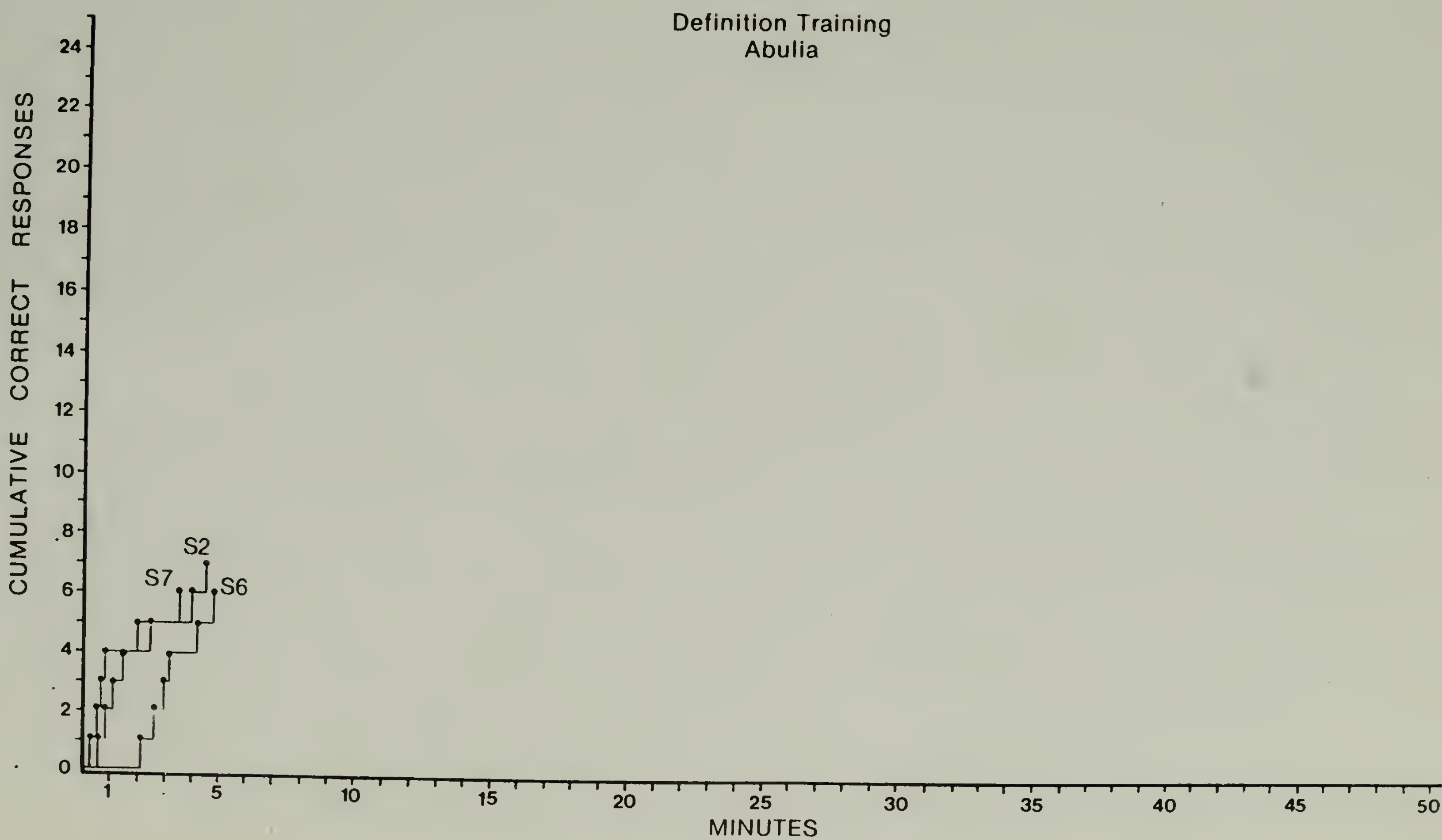
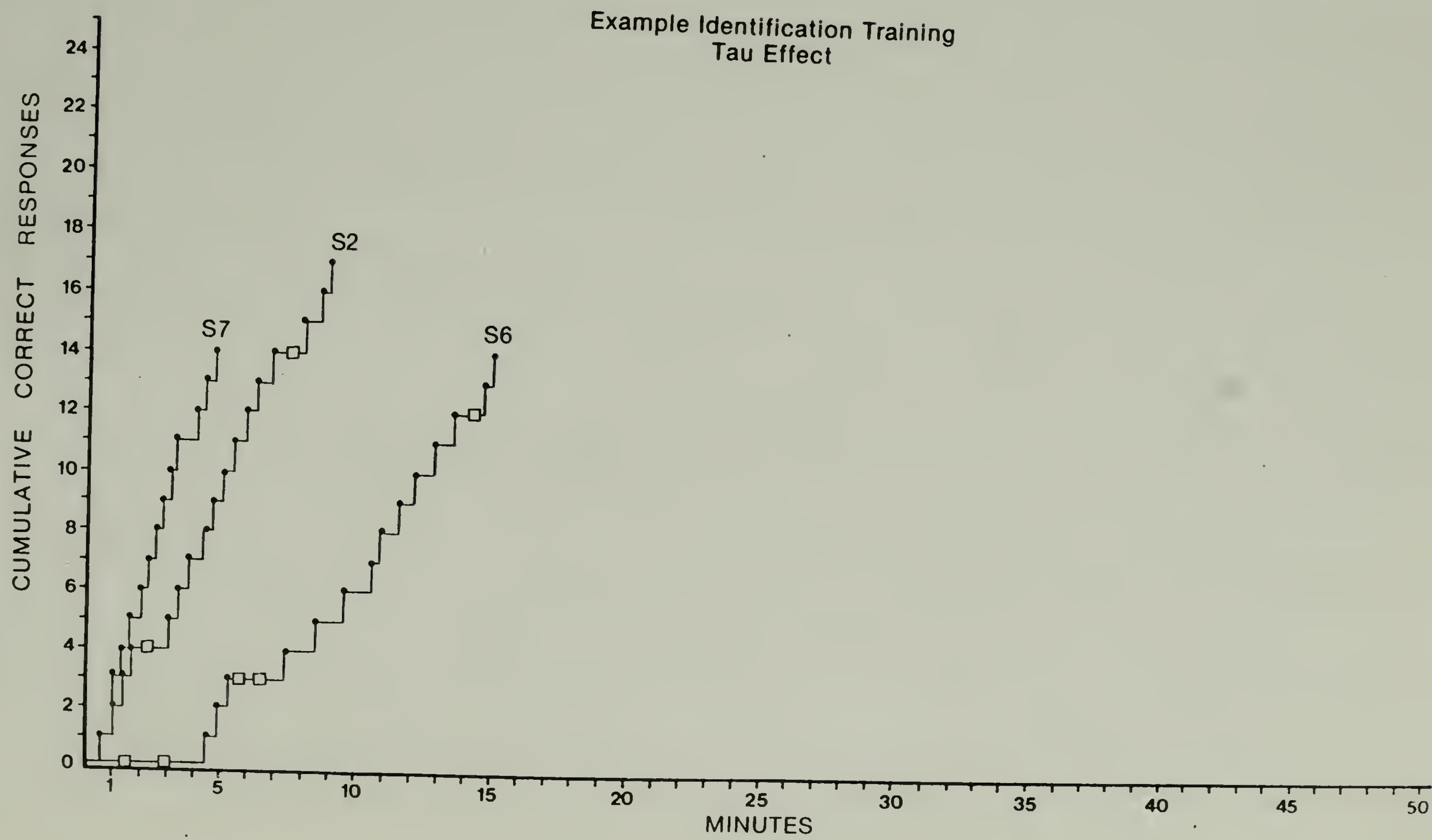
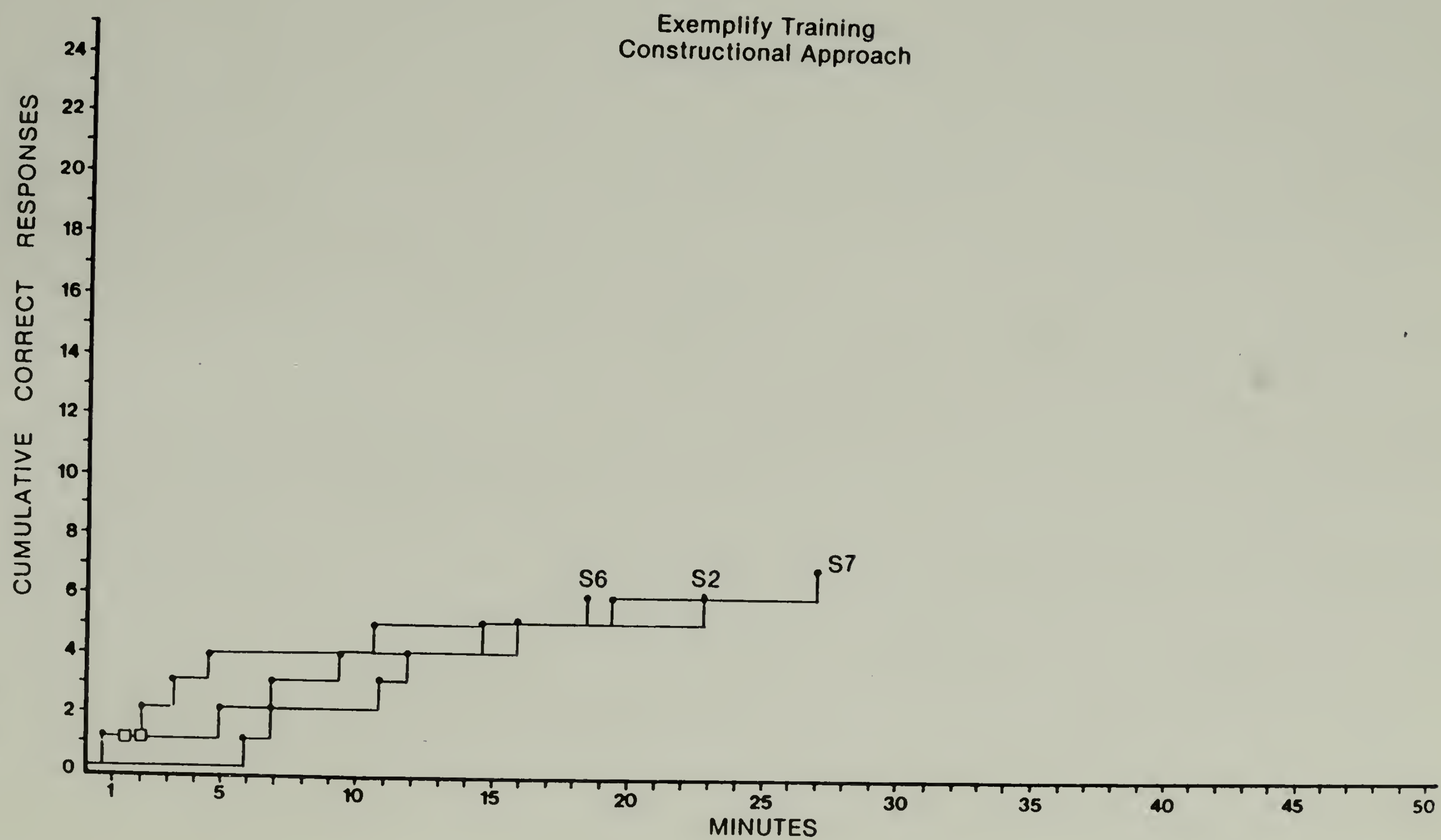


Figure 3. Cumulative frequency of correct study performance.


Example Identification Training
Tau Effect
CUMULATIVE CORRECT RESPONSES
0
2
4
6
8
10
12
14
16
18
20
22
24
S7
S2
S6
1
5
10
15
20
25
30
35
40
45
50
MINUTES


Exemplify Training
Constructional Approach
CUMULATIVE CORRECT RESPONSES
0
2
4
6
8
10
12
14
16
18
20
22
24
MINUTES
1
5
10
15
20
25
30
35
40
45
50
S6
S2
S7

Second, Figures 1-3 illustrate more rapid rates of correct responding on example identification questions than on either definition questions or exemplify questions. Seven of the nine subjects' response rates on example identification questions were higher than the rates on definition questions (Subjects 1-5; 7&8). Eight of the subjects responded faster on example identification questions than on exemplify questions (Subjects 1-8). A planned comparison between types of study programs revealed a significant difference between exid. and the other programs, $F(1, 12) = 6.82$, $p < .025$.

These data were separated into errors and duration of study program and other differences were revealed. The example identification program took less time to complete than the exemplify program for seven subjects (Subjects 2-8). However, this difference was not significant. Conversely, eight subjects made more errors on example identification questions than either exemplify or definition questions (Subjects 1-7, 9). This difference was found to be significant, $F(1, 12) = 7.42$, $p < .025$.

Third, Figures 1-3 demonstrate more rapid rates correct responding on definition tasks than on exemplify tasks. Correct definition rates were higher than exemplify rates for six subjects (Subjects 1-2; 4-7). Error analyses revealed that the performance of six subjects was errorless on definition questions and one subject made a single error. Analyses of duration showed that the definition program took less time to complete than the exemplify program for six subjects (Subjects 1, 2, 5-7, & 9). Systematic differences in duration between example identification and definition programs were not found; four subjects finished the example identification program faster while five subjects completed the definition program more rapidly. In addition, planned comparisons of study duration did not reveal any significant differences between study programs.

Fourth, exemplify performances were errorless for seven of the subjects (1, 2, 5-9). The other two subjects made one incorrect response each.

In summary, these analyses revealed a number of relations among concepts and among study programs. Constructional approach was shown to be more difficult than either abulia or tau effect for most subjects. Differences in study programs varied with the different dependent variables: Example identification programs produced higher rates of correct responding. Definition and exemplify performance was virtually errorless and exemplify programs took much longer for most subjects to complete. Table 6 presents each of these relations with the proportion of subjects whose performance corresponded to each specified relation.

- - - - - - - - - - - - - - -

SEE TABLE 6 PAGE 70

- - - - - - - - - - - - - - -

Further group analyses of each measure of study performance were calculated to statistically substantiate the intrasubject analyses. In addition, the effect of order of presenting the study programs was analyzed and interactions between order and other factors were estimated.

First, a three way repeated measure, Latin square analysis of variance (ANOVA) was conducted for each dependent measure. Specifically, a 3 (order of study program) x 3 (concept) x 3 (study program) ANOVA with order as the Latin square factor, concepts as a repeated measures factor and study program as the within square factor were calculated for each dependent measure. An arc sin transformation of proportions was used for all percent correct data throughout the study. These transformations were conducted because the variances obtained from proportions are always systematically related to the mean (Myers, 1979). Therefore, arc sin transformations were used to stabilize the variance in order to assume homogeneity of variance.

Significant effects of study program order, and the interactions between the order effect and the effect of concept were not found for any dependent measure of study performance. For example, Table 7 presents the source data for the percentage of correct responses made during study

TABLE 6

SUMMARY OF INTRASUBJECT ANALYSES OF STUDY PERFORMANCE. ALL RELATIONS REPRODUCED FOR 6 OR MORE SUBJECTS ARE PRESENTED.

Independent Variable / Dependent Measure / Specific Relation*	Proportion of Cases (N=9)
Study Program	
Rate of Correct Responses	
Example ID. Program > Define Program	.78
Example ID Program > Exemplify	.89
Define Program Exemplify	.67
Number of Errors	
Example ID. Program > Define Program	.89
Example ID. Program > Exemplify	1.00
Define Program = Exemplify Program	.67
Study Duration	
Example ID. Program < Exemplify	.78
Define Program < Exemplify Program	.67
Concept	
Rate of Correct Response	
Constructional Approach < Abulia	.89
Constructional Approach < Tau Effect	.89
Study Duration	
Constructional Approach < Abulia	1.00
Constructional Approach < Tau Effect	.89

* Planned comparisons of statistical differences are presented in Table 8.

performance. The absence of order effects and an order by concept interaction indicated that tests for the simple effects of concept and study program were unbiased.

- - - - - - - - - - - - - - -

SEE TABLE 7, PAGE 72

- - - - - - - - - - - - - - -

Two of the three accuracy measures revealed a significant main effect of study program. Analyses of study program effects were obtained by partitioning the order of study program by concept interaction into study program and residual effects. An F test of this partition for the number of errors yielded a significant effect of study program, $F(2, 12) = 4.12$, $p < .05$.

The ANOVA for correct responses per minute also revealed a significant main effect of study program, $F(2, 12) = 4.27$, $p < .05$. As reported earlier the planned comparison of the different study programs yielded significantly higher rates of responding on the example identification programs than on both definition and exemplify programs.

The ANOVA for percent correct study performance did not result in a significant effect of study program, $F(2, 12) = 3.67$, $p > .05$.

All three analyses of correct performance resulted in a nonsignificant affect of concept. For number of errors the ANOVA resulted in an $F(2, 12) = .99$, $p > .05$. For correct

TABLE 7

ANALYSIS OF VARIANCE FOR PERCENT CORRECT PERFORMANCE ON STUDY PROGRAMS

Source	Sum of Squares	DF	Mean Squares	F
Mean	209264.037	1	209264.037	676.58
Order (A)	64.518	2	32.259	.10
Error	1855.777	6	309.296	
Concept (B)	332.074	2	166.037	.99
AB				
Program (C)	1224.520	2	612.20	3.67
Residual	207.48	2	153.74	.92
Error	2005.555	12	167.129	

responses per minute the ANOVA yielded an $F(2, 12) = 1.75$, $p > .05$.

An ANOVA of study duration yielded significant relations that were not revealed by accuracy measures. The effect of study program was not significant, $F(2, 12) = 1.31$, $p > .05$. However, the effect of concept was significant, $F(2, 12) = 13.98$, $p < .05$. Since a planned comparison between _constructional approach_ and both _abulia_ and _tau effect_ was significant, it appears that the constructional approach takes significantly more time to complete than either of the other concepts.

In summary, group analyses of the number of study errors, the percentage of correct study answer, the time to complete the study program and the rate of correct responding during study trials revealed a number of statistically significant relations. Table 8 presents each of these.

- - - - - - - - - - - - - -

SEE TABLE 8, PAGE 74

- - - - - - - - - - - - - -

Test performance. Total test performance was analyzed with three kinds of data: rate of correct responding, duration of the test, and percent correct. The transfer test consisted of definition, exemplify, example identification and combination tasks. Intrasubject analyses of these data revealed several relations. Systematic differences on test performance were found to be related to both type of study

TABLE 8

SUMMARY OF SIGNIFICANT RELATIONS BETWEEN STUDY PROGRAM AND DEPENDENT MEASURES OF STUDY PERFORMANCE, AND CONCEPTS AND DEPENDENT MEASURES OF STUDY PERFORMANCE

Independent Variable Dependent Measure	F	Planned Comparison	F
Study Program			
Rate of Correct Responses	4.27*	Exam Id. > Def. or Exem.	6.82**
Number of Errors	4.12*	Exam. Id. > Def. or Exem.	7.42**
Study Duration	-	-	-
Percent Correct	-	-	-
Concept			
Rate of Correct Responses	-	-	-
Number of Errors	-	-	-
Percent Correct	-	-	-
Study Duration	***13.98	Con. App. > Abulia or T.E.	27.79***

* indicates α level of .05

** indicates α level of .025

*** indicates α level of .01

program and concept.

Figures 4-6 present the correct responses per minute for the nine subjects on each transfer test after each kind of study program. Subjects were grouped on the individual graphs according to the combination of concept and program that they received. The same graphing convertions were used as those used for Figures 1-3.

- - - - - - - - - - - - - - - - - -

SEE FIGURES 4-6, PAGES 76-84

- - - - - - - - - - - - - - - - - -

First, Figures 4-6 illustrate more rapid rates of correct responding after example identification programs than after definition programs. Six of the nine subjects' rates were higher after example identification training than after definition training (Subjects 2-4; 6-8). Comparisons of correct responding after exemplify programs and example identification programs were less systematic. Five subjects had higher rates after example identification programs and four had higher rates after exemplify programs. However, a planned comparison between example identification and the other two programs revealed a significant difference, $F(1, 12) = 12.77$, $p < .01$.

Second, Figures 4-6 indicate that subjects took more time to complete tests after exemplify programs than after either definition or example identification programs. Six subjects spent more time on the transfer test after exemplify

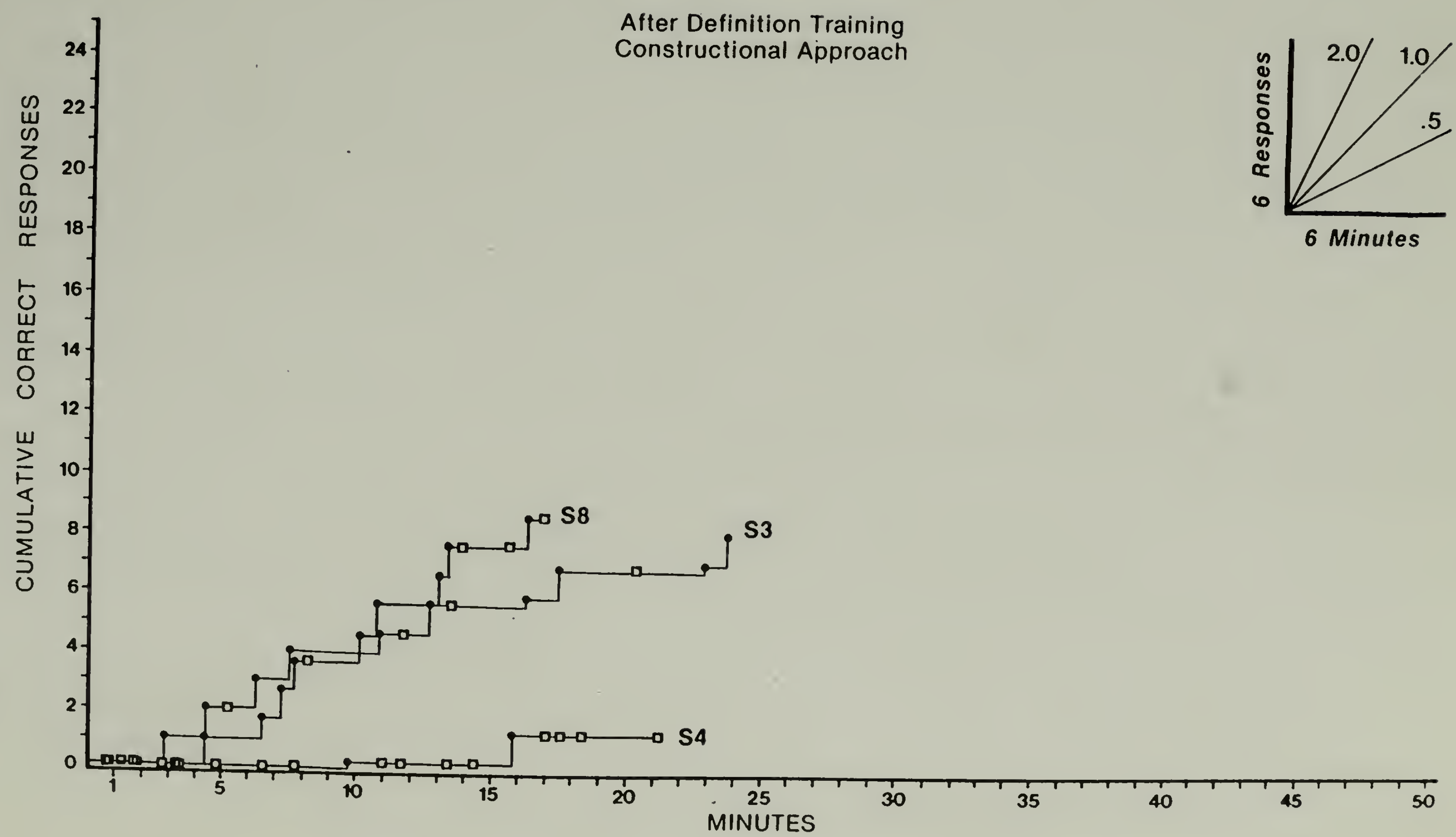
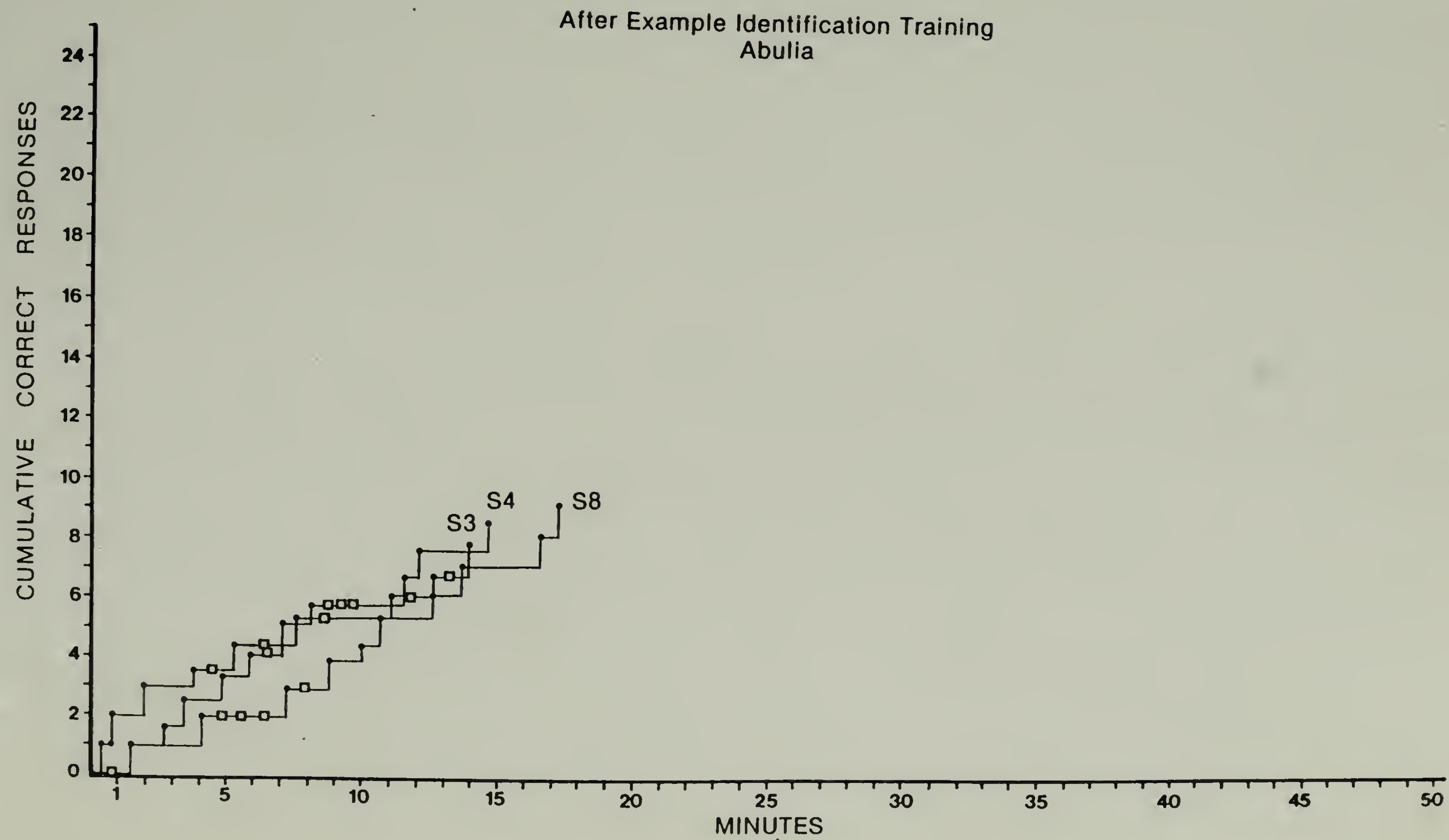
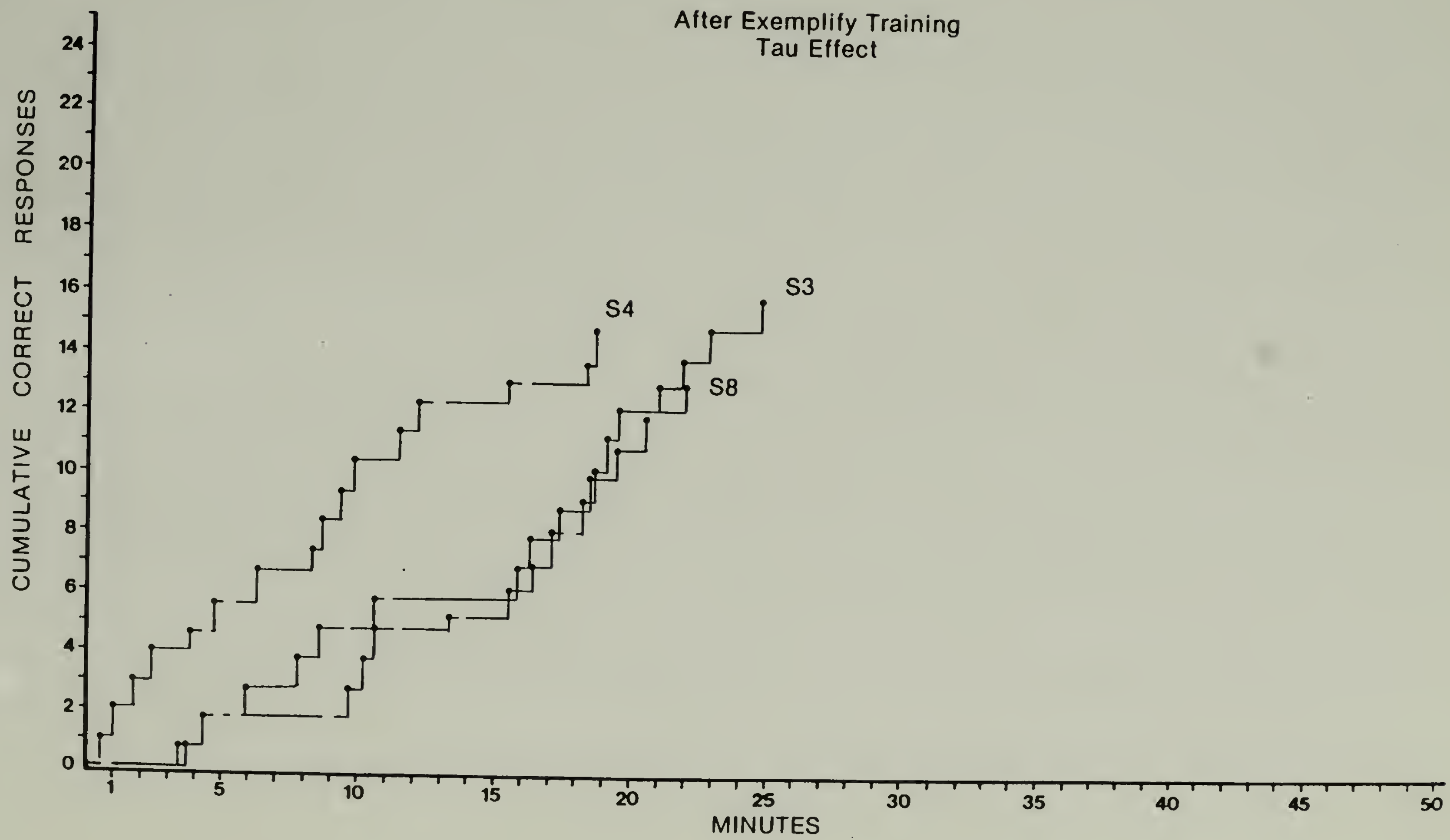


Figure 4. Cumulative frequency of correct test performance.



After Example Identification Training
Abulia
CUMULATIVE CORRECT RESPONSES
0
2
4
6
8
10
12
14
16
18
20
22
24
1
5
10
15
20
25
30
35
40
45
50
MINUTES
S3
S4
S8


After Exemplify Training
Tau Effect
CUMULATIVE CORRECT RESPONSES
MINUTES
S4
S3
S8
0
2
4
6
8
10
12
14
16
18
20
22
24
1
5
10
15
20
25
30
35
40
45
50

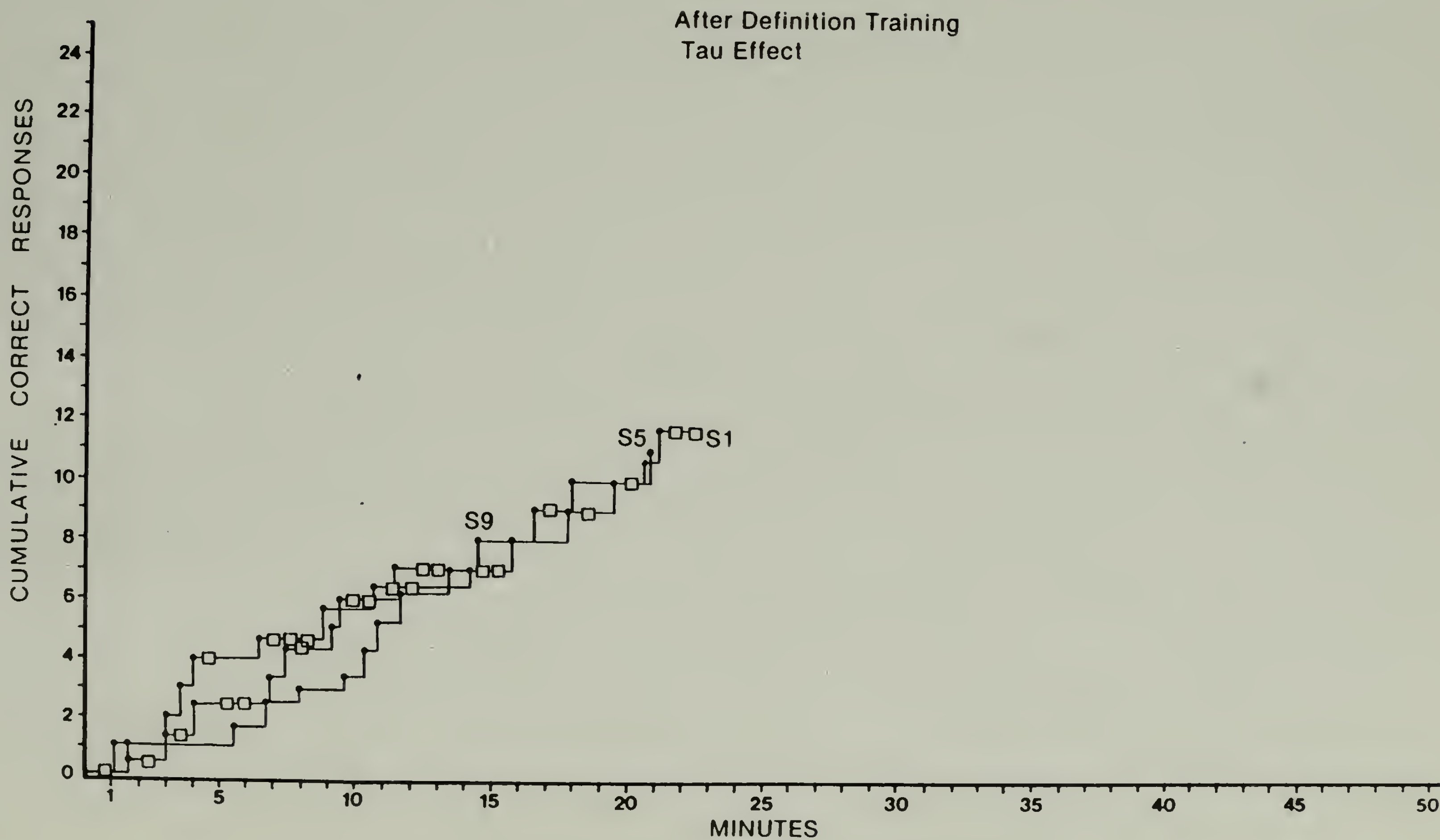
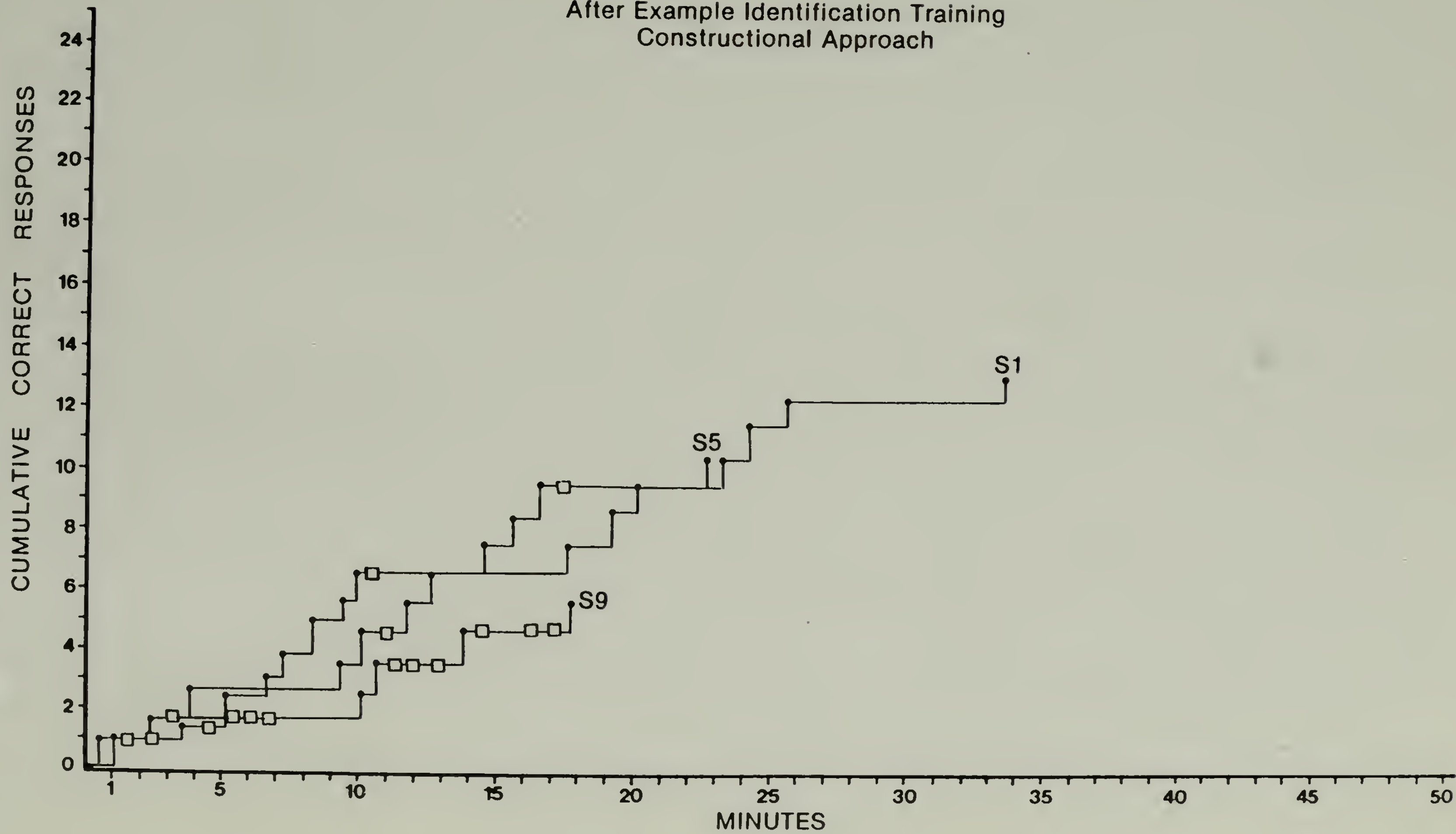
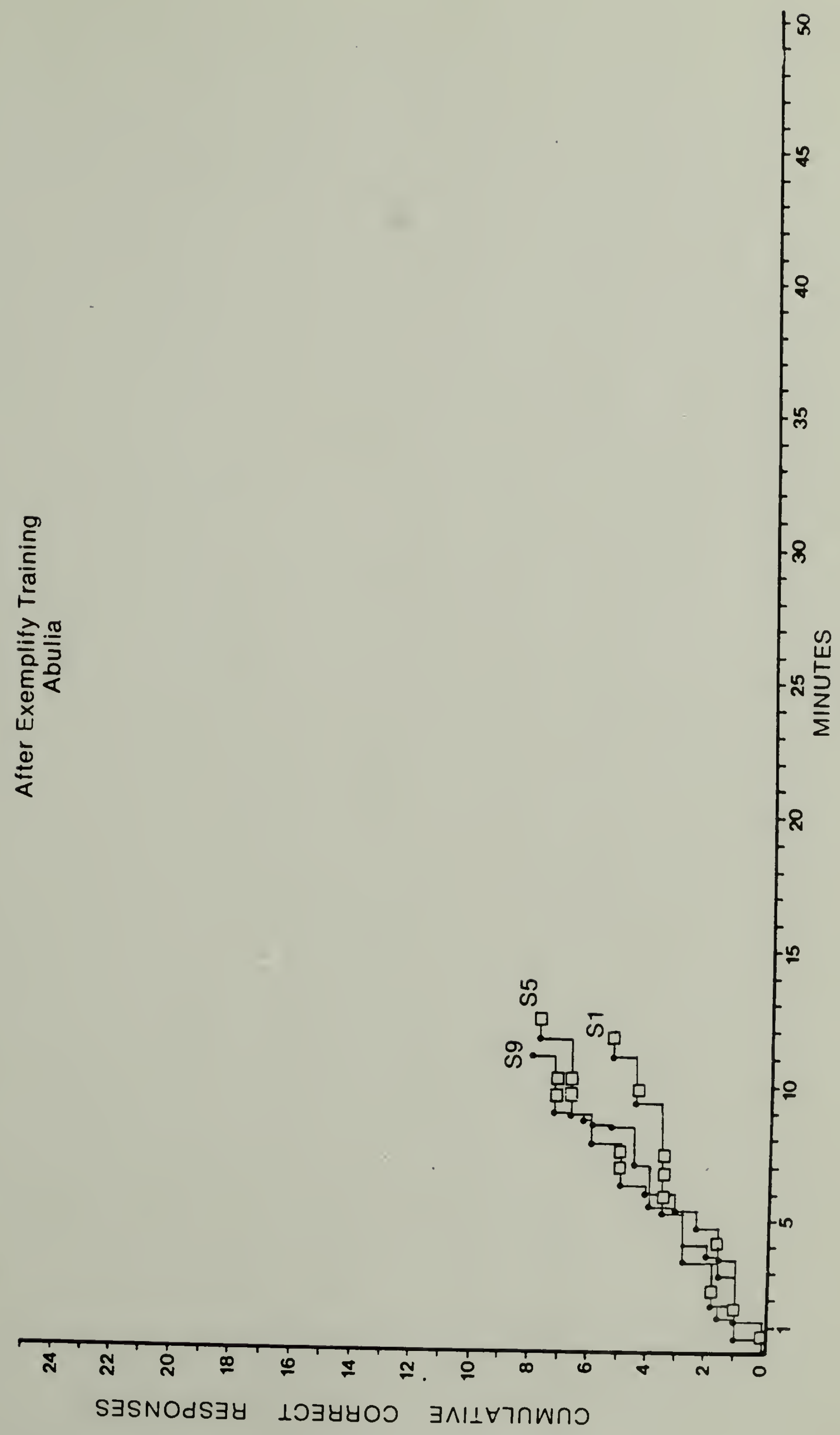


Figure 5. Cumulative frequency of correct test performance.

After Example Identification Training
Constructional Approach




After Exemplify Training
Abulia
CUMULATIVE CORRECT RESPONSES
MINUTES
S9
S5
S1

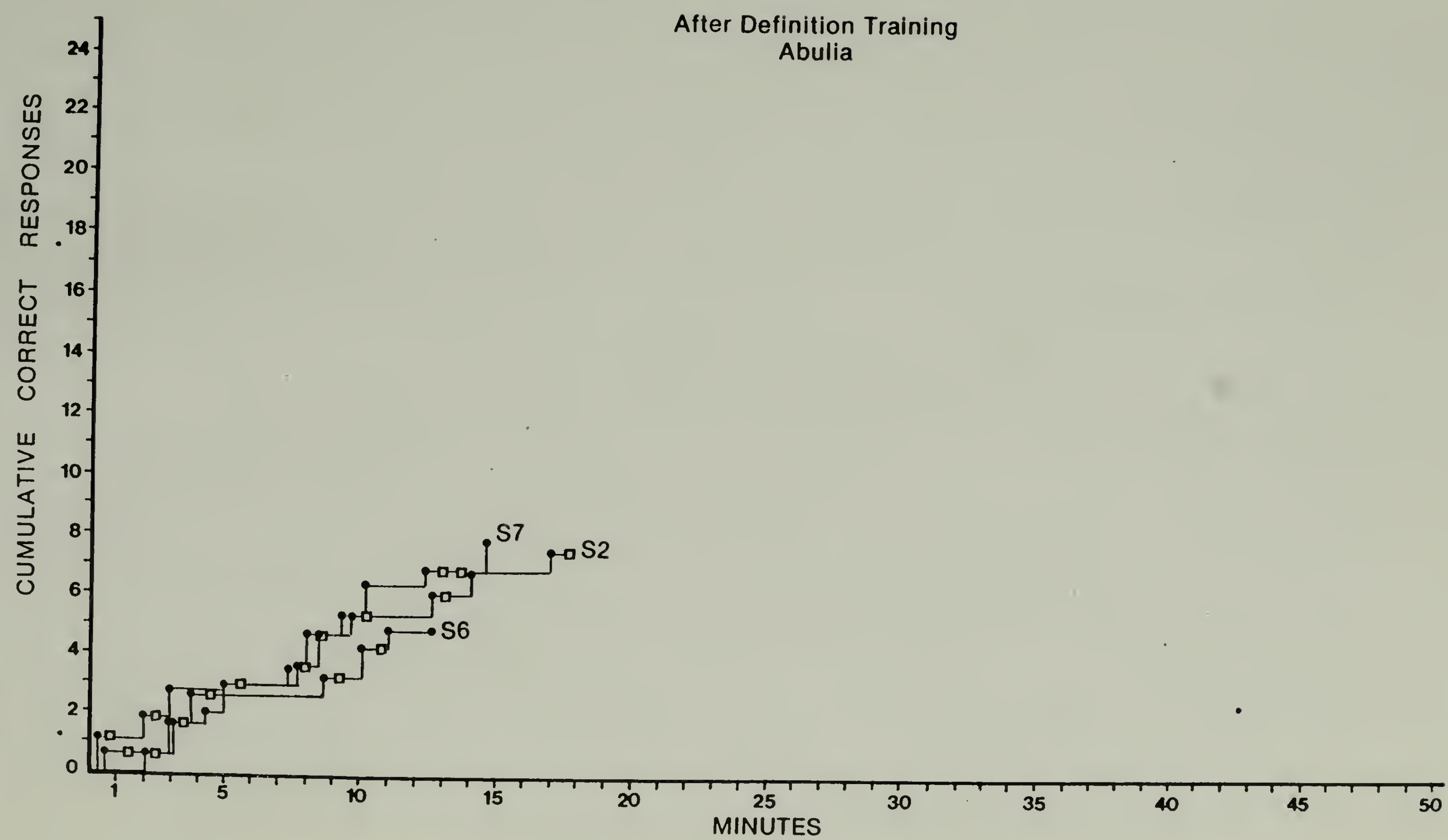
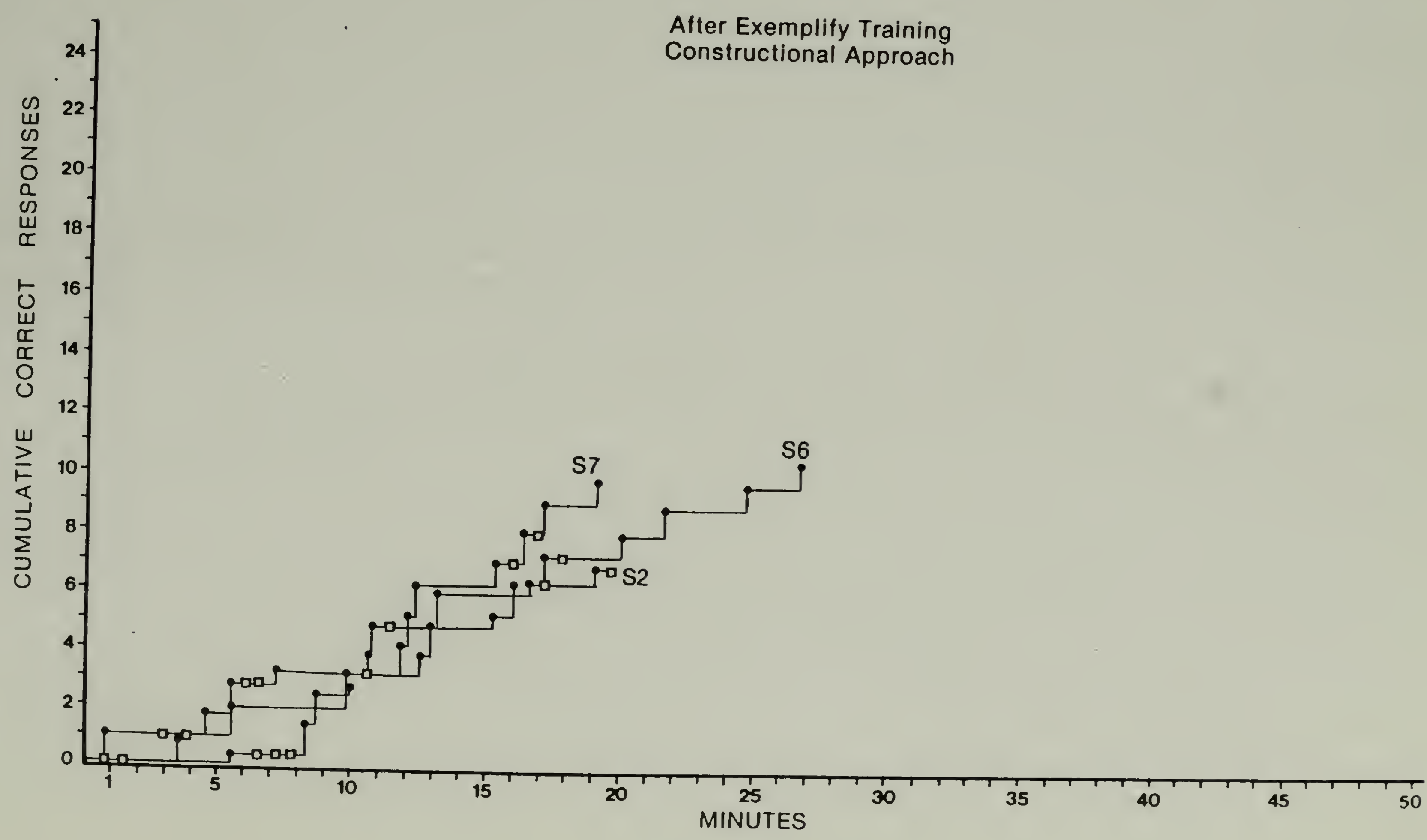
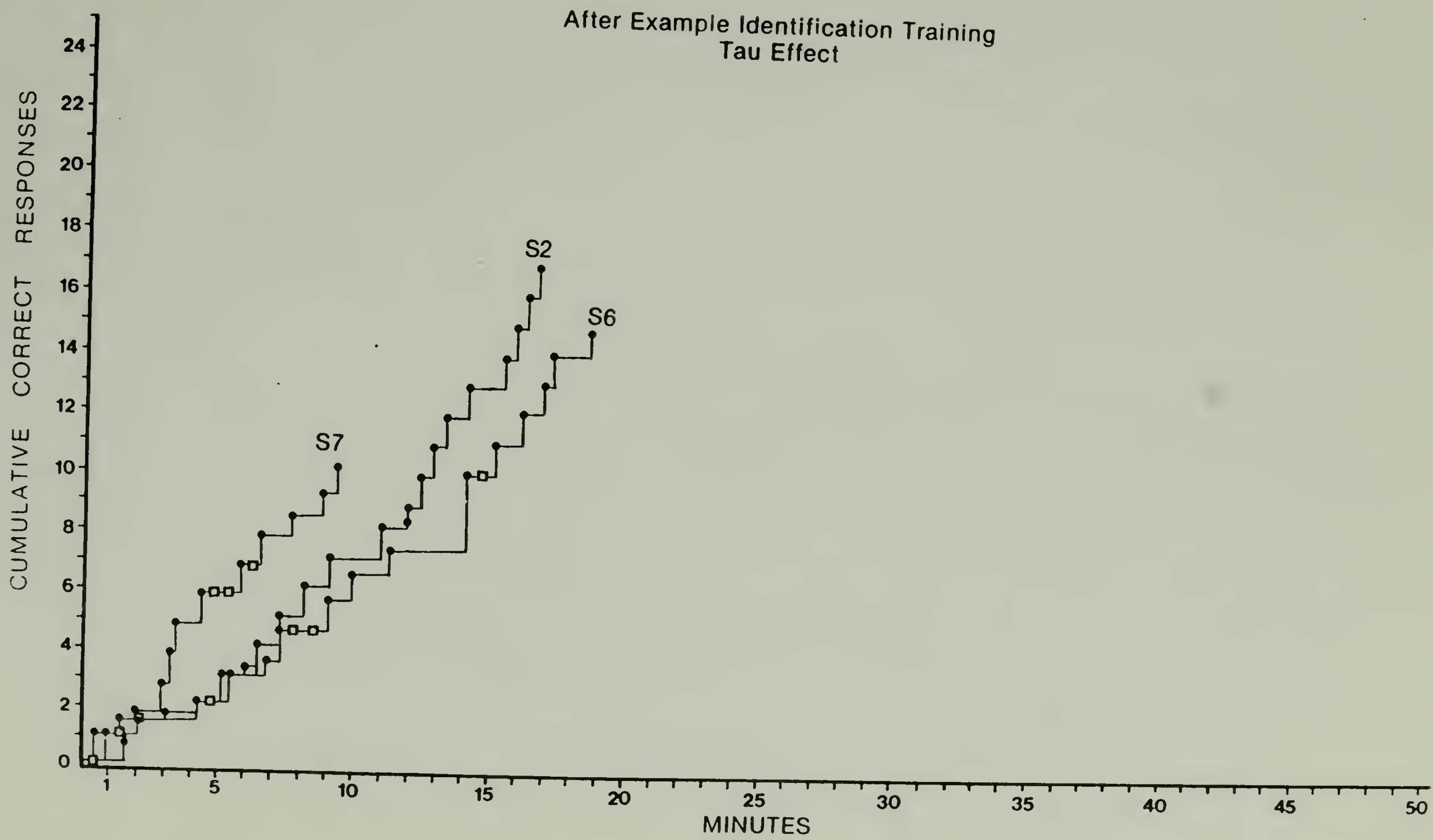


Figure 6. Cumulative frequency of correct test performance.


After Exemplify Training
Constructional Approach
CUMULATIVE CORRECT RESPONSES
24
22
20
18
16
14
12
10
8
6
4
2
0
1
5
10
15
20
25
30
35
40
45
50
MINUTES
S7
S6
S2


After Example Identification Training
Tau Effect
CUMULATIVE CORRECT RESPONSES
MINUTES
S7
S2
S6
0
2
4
6
8
10
12
14
16
18
20
22
24
1
5
10
15
20
25
30
35
40
45
50

training than after definition training (Subjects 2-4; 6-8). Six subjects also spent more time on the transfer tests after exemplify training than after example identification training (Subjects 2-4, 6-8). Differences in test duration were less systematic when definition and example identification programs were compared. Five subjects completed the tests more rapidly after definition training and four subjects completed the tests faster after example identification training. None of the differences in test duration were significant.

Third, these data revealed slower rates of correct responding on the constructional approach tests than on either the abulia or the tau effect tests. The rates of eight subjects were lower for constructional approach than for abulia (Subjects 2-9). Nine subjects' rates were lower for constructional approach than for tau effect (Subjects 1-9). In addition, six subjects' correct response rates were lower for abulia tests than for tau effect tests. Planned comparisons between concepts revealed that all of these differences were significant. Constructional approach was found to be significantly different from both abulia and tau effect, $F\ (1, 12) = 30.16$, $p < .01$. Tau effect was also found to be significantly different from abulia, $F\ (1, 12) = 12.12$, $p < .01$.

Fourth, Figures 4-6 demonstrate that subjects spent more time on tests for constructional approach than either *abulia* or *tau effect* tests. Nine subjects took longer to complete the tests for *constructional approach* than for *abulia*. Seven subjects took longer to complete the test for *constructional approach* than for *tau effect*. (Subjects 1-2; 4-7, 9). In addition, eight subjects took longer to complete the tau effect tests than the abulia tests (Subjects 1-6; 8, 9). Both of these differences were also found to be significant through planned comparisons. *Constructional approach* was significantly lower than both *abulia* and *tau effect* $F\ (1, 12) = 12.35$, $p < .01$ and *tau effect* was significantly lower than *abulia* $F\ (1, 12) = 5.10$, $p < .025$.

Figure 7 presents each subjects percentage of correct answers on each test. The data were grouped by concept and by type of training. The symbols C, A and T were used to represent *constructional approach*, *tau effect* and *abulia* respectively. These percent correct revealed the same kinds of systematic concept differences as the rate data. However, percent correct data revealed slightly different relations between study programs.

First, percent correct performance on *constructional approach* was lower than on abulia for six subjects (Subjects 2-5; 8, 9). Percent correct performance on the *constructional approach* was also lower than tau effect for seven subjects (Subjects 2-5; 7-9). In addition, percent correct performance on *tau effect* was also higher than performance on *abulia* for six subjects (Subjects 1-4; 6, 7). However, none of these differences was significant.

Second, percent correct performance after example identification training was higher than after definition training for seven subjects (Subjects 1-4; 6-9). After exemplification training percentages were also higher than after definition training in seven cases (Subjects 1; 3-8). These differences were significant. A planned comparison between definition and both example identification and exemplify revealed that performance after definition training was significantly lower, $F(1, 12) = 5.10$, $p < .05$. The differences between percent correct performance after example identification and exemplify training were less systematic.

- - - - - - - - - - - - - - -

SEE FIGURE 7, PAGE 88

- - - - - - - - - - - - - - -

Subjects 3-6 and 9 scored higher after exemplify programs, subjects 1, 2, 7 and 8 scored higher after example identification training.

Test performances were also analyzed by separating tests into extension and transfer performance. Thus, correct rate of responding and percent correct measures were analyzed for test items that were novel, but of the same type as those received in training, extension tasks, and item types for categories not trained, transfer tasks.

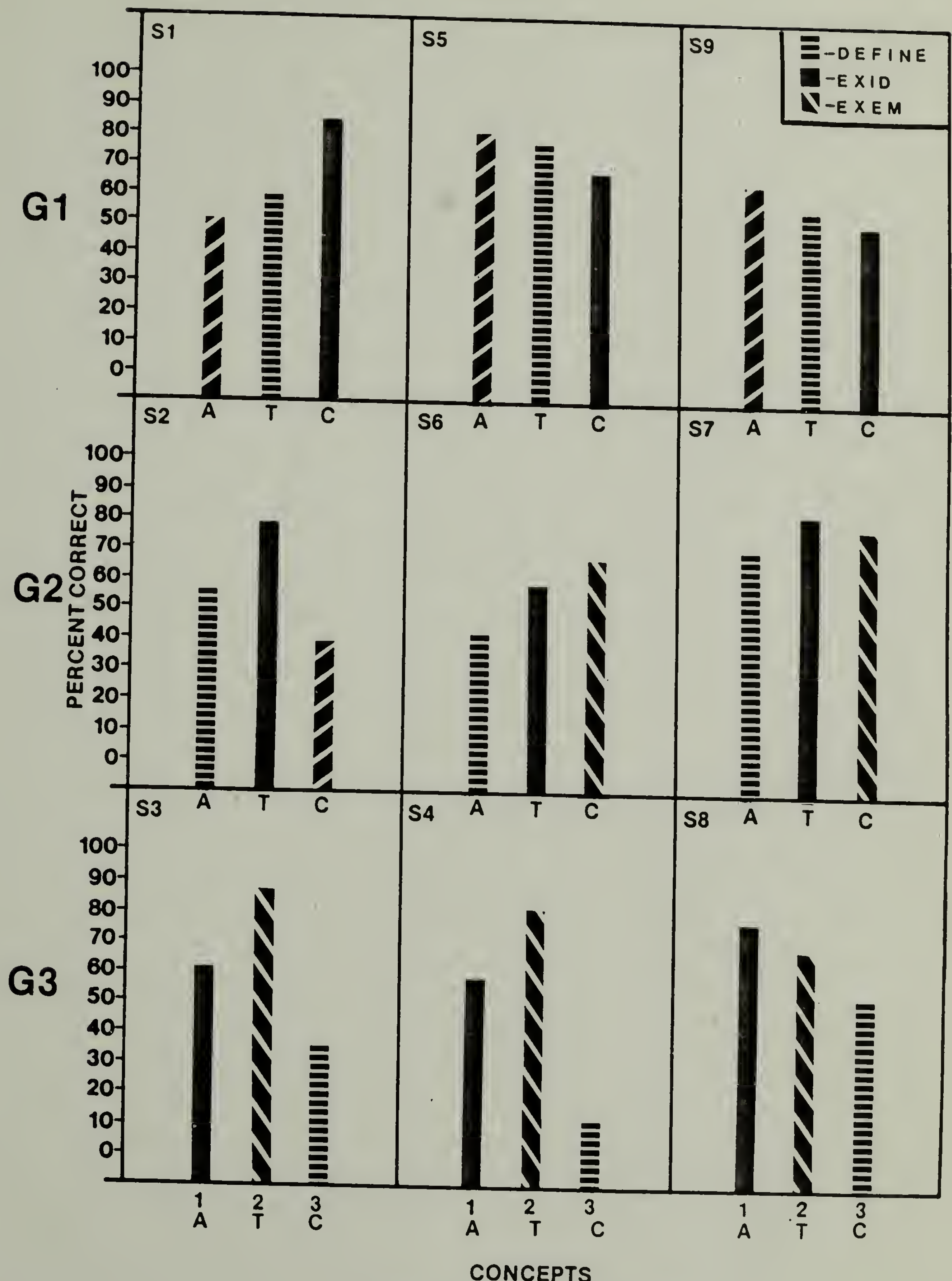


Figure 7. Total test performance.

Systematic differences among extension and transfer performances were found to be related to both type of training and concept.

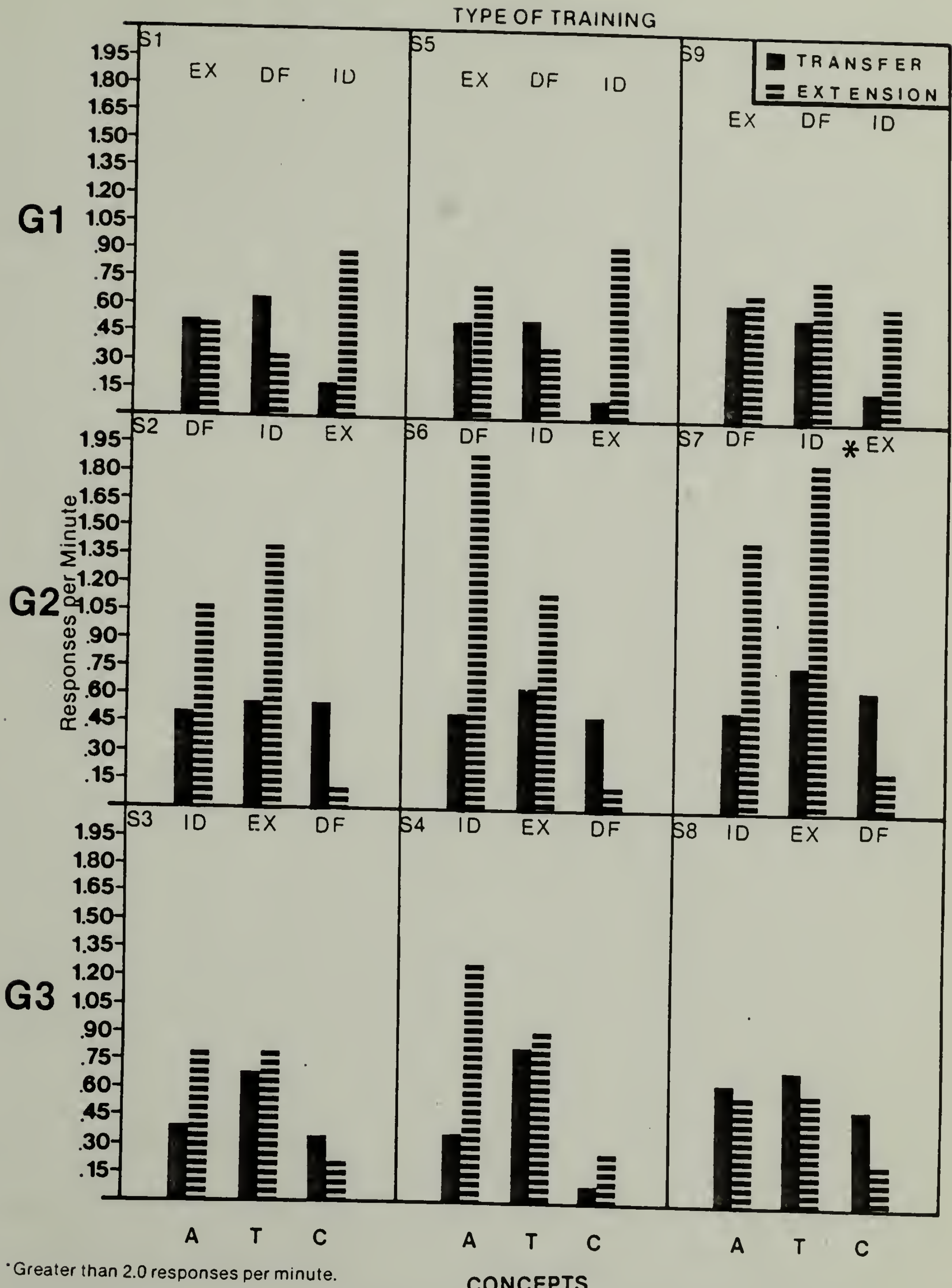
Figure 8 presents the rate of correct responding for both extension and transfer tasks for all nine subjects. The data were grouped by concept and by type of training. The symbols C, A, and T represent constructional approach, abulia and tau effect respectively. The symbols Ex., Id., and Df. represent exemplify, example identification and definition training respectively. Subjects were grouped according to the combination of concept and study program that they received.

- - - - - - - - - - - - - - -

SEE FIGURE 8, PAGE 90

- - - - - - - - - - - - - - -

First, extension performance was higher than transfer performance for most subjects. Sixteen of the twenty-seven intrasubject comparisons revealed this difference. In five of the eleven exceptions, the subjects were tested on constructional approach and extension performance was very low. By refering back to Figures 1-3, it can be seen that in all five of these cases the subjects had low rates of correct responding during study trials (Subjects 2, 6 and 7 on exemplify training; subjects 3 and 8 on definition training). A planned comparison between extension and transfer



*Greater than 2.0 responses per minute.

Figure 8. Correct responses per minute on transfer and extension tasks.

scores revealed significantly higher extension rates than transfer rates, t (52) = 2.23, $p < .05$.

Second, extension performances after example identification training were higher than after definition and exemplify training. Extension performance after example identification training was higher than after exemplify training for six subjects (Subjects 1, 2, 4-7). Extension performance after example identification training was higher than after definition training for the same six subjects. A planned comparison between example identification training and the other two types of training revealed significantly higher performance after example identification training, F (1, 12) = 11.76, $p < .01$. Thus, rate of correct extension performance maintained the same relation between study programs as found during study trials.

Third, extension performance on constructional approach tests were lower than extension performance on either abulia or tau effect tests. Subjects 2-4, 6-9 had lower correct response rates on constructional approach than on abulia and tau effect. A planned comparison between concepts revealed that this difference was also significant F (1, 12) = 12.12, $p < .01$. Thus, the difference between concepts found during study trials was maintained on the test.

Fourth, the effect of training on transfer performance varied across subjects. However, for six subjects performance after exemplify training was higher than performance after example identification training (Subjects 1, 3-5, 8 and 9). In addition, for six subjects the rate of correct performance after example identification training was higher than performance after example identification training (Subjects 1, 3, 4, 5, 8 and 9). In addition, for six subjects rate of correct performance after example identification training was higher than performance after definition training (Subjects 2-4, 6-8). However, planned comparisons between the effects of training revealed that only the difference between exemplify training and the other two types of training was significant, $F (1, 12) = 16.56$, $p < .01$.

Fifth, the transfer data indicated differences between concepts. Six of the subjects had lower transfer performance on constructional approach tests than on both abulia and tau effect tests (Subjects 2-5, 8 and 9). This difference was significant, $F (1, 12) = 29.37$, $p < .01$. Eight subjects had higher transfer performance on tau effect tests than on abulia tests (Subjects 1-8), yet this difference was not significant.

Figure 9 presents the percent correct performance for each subject on both extension and transfer tasks. Graphing conventions similar to Figure 9 were used. These data revealed a number of systematic relations.

- - - - - - - - - - - - - - -

SEE FIGURE 9, PAGE 93

- - - - - - - - - - - - - - -

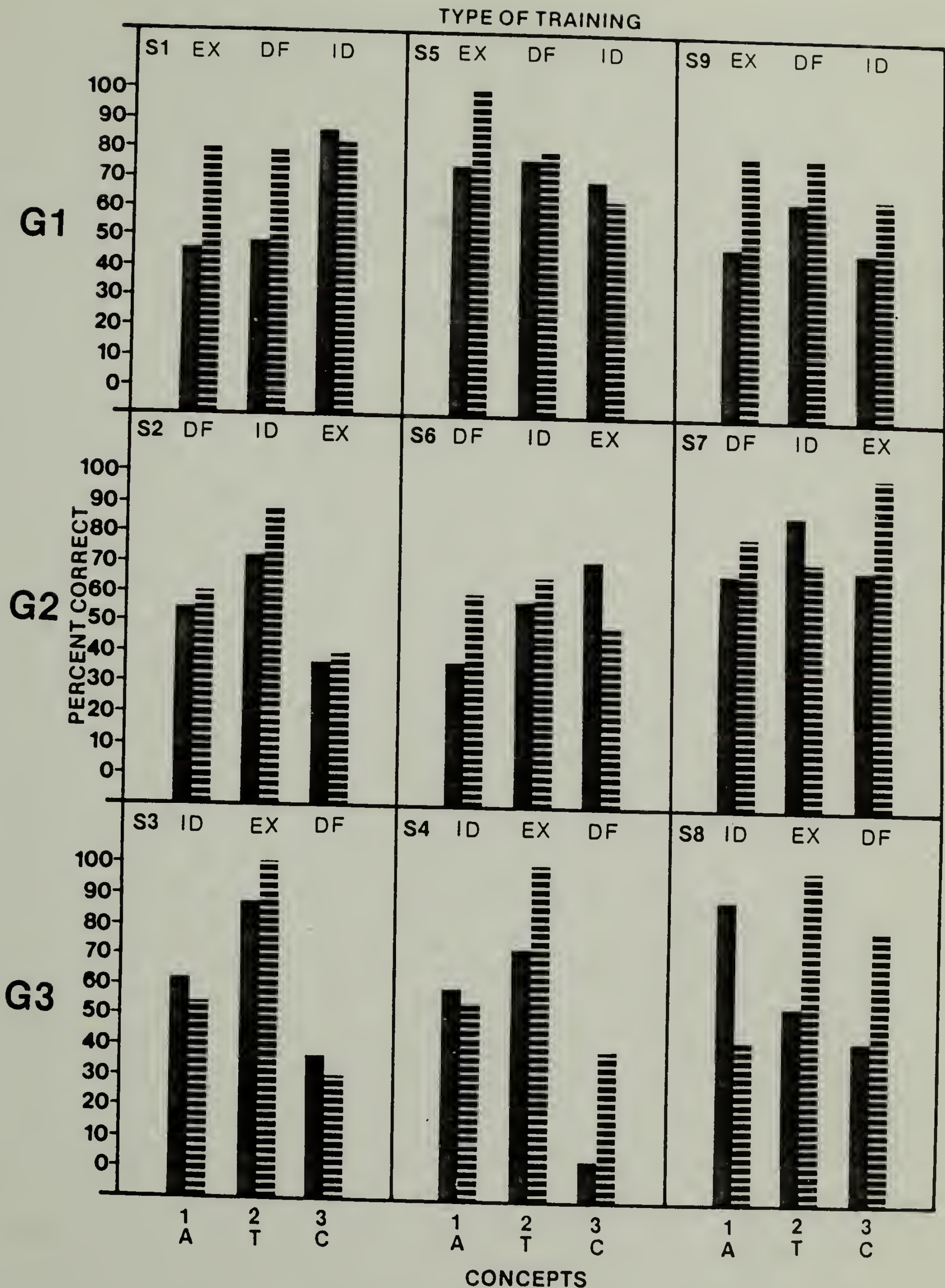


Figure 9. Percent correct on transfer and extension tasks.

First, extension performance was generally more accurate than transfer perforamnce. On twenty of the twenty-seven tests, subjects' extension performance was higher than their transfer performance. Four of the eight exceptions occurred with the constructional approach. A comparison between extension and transfer scores revealed that extension scores were significantly higher than transfer, $t\ (52) = 1.71,\ p < .05$.

Second, extension performance after exemplify training was higher than after other types of training. The extension performance of six subjects (3, 4, 5, 7, 8, & 9) was higher after exemplify training than after example identification training. The extension performance of five subjects (3-5, 7, 8) was higher after exemplify training than after definition training. Two additional subjects had equal performance after exemplify and definition training. These differences were significant. A planned comparison between performance after exemplify training and performance after both the other types of training revealed a significant difference, $F\ (1,\ 12) = 5.97,\ p < .05$. Thus, on percent correct measures the relation between different types of training was maintained on the test.

Third, extension performance was systematically affected by concept. Extension performance on the _tau effect_ tests was higher than on the _constructional approach_ tests for seven subjects (Subjects 2-6, 8 and 9). Extension performance on the _tau effect_ tests was higher than the performance on the abulia tests for six subjects (Subjects 2-4, 6, 8 and 9). Extension performance on the abulia tests was higher than performance on the _constructional approach_ tests for six subjects (Subjects 2-6 and 9). Planned comparisons between concepts revealed that accuracy was significantly lower for _constructional approach_ for either _abulia_ or _tau effect_, $F\ (1, 12) = 4.49$, $p < .05$. However, _tau effect_ was not significantly different from _abulia_.

Fourth transfer performance after example identification training was higher than after definition training for seven subjects (Subjects 1-4, 608). However, differences between example identification and exemplify training, and between exemplify training and definition training were less systematic. In addition, none of these relations was significant.

Fifth, transfer performance on tests for _tau effect_ was higher than performance for either _abulia_ or _constructional approach_. Transfer performance of eight subjects on tau effect was higher than on abulia (Subjects 1-7 and 0). In addition, transfer performance on the abulia tests was higher than on constructional approach tests for six subjects (subjects 2-5, 8 and 9). However, none of these relations was significant.

In summary, comparisons between study programs and concepts revealed a number of systematic relations, some of these were not statistically significant. Others occurred for most subjects and were large enough to be found significant. Table 9 presents all of the intrasubject relations that occurred for six or more subjects within a group of nine. In addition, extension versus transfer performance relations are also indicated (n = 27). The proportion of subjects whose data corresponded to the relation is indicated.

- - - - - - - - - - - - - - -

SEE TABLE 9, PAGES 97-98

- - - - - - - - - - - - - - -

Further group analyses of rate of correct responding, percent correct and total test duration were conducted to statistically substantiate the analyses described above. In addition, the effect of order of study program was analyzed and the interaction between order and concepts was estimated.

As with the group analyses of study performance, a 3 way repeated measures, latin square ANOVA was calculated for each dependent variable. Order of study program was the latin square factor, concept was the repeated factor and study program was the within square factor. Each factor had three levels. Significant main effects were found and interactions were estimated from a number of analyses. The following details these results.

TABLE 9

SUMMARY OF INTRASUBJECT ANALYSES OF TEST PERFORMANCE. ALL RELATIONS REPRODUCED FOR 6 OR MORE SUBJECTS ARE PRESENTED

Independent Variable Dependent Variable Specific Relation	Proportion of Cases (N=9 unless indicated)
Study Program	
Rate of Correct Responses	
Extension Tasks $>$ Transfer Tasks	.67 (n=27)
Example Id. $>$ Definition on Total Test	.67
Example Id. $>$ Definition on Extension	.67
Example Id. $>$ Exemplify on Extension	.67
Example Id. $>$ Definition on Transfer	.67
Exemplify $>$ Example Id. on Transfer	.67
Exemplify $>$ Definition on Transfer	.67
Percent Correct	
Extension Tasks $>$ Transfer Tasks	.74 (n=27)
Example Id. $>$ Definition on Total Test	.78
Exemplify $>$ Definition on Total Test	.78
Exemplify $>$ Example Id. on Extension	.67
Exemplify $>$ Definition on Extension	.78
Example Id. $>$ Definition on Transfer	.78
Test Duration	
Exemplify $>$ Definition	.67
Exemplify $>$ Example Id.	.67
Concept	
Rate of Correct Responses	
Con. App. $<$ Tau Effect on Total Test	1.00
Con. App. $<$ Abulia on Total Test	.89
Abulia $<$ Tau Effect on Total Test	.67
Con. App. $<$ Tau Effect on Extension Item	
Con. App. $<$ Abulia on Extension Items	.78
Con. App. $<$ Tau Effect on Transfer Items	.78
Con. App. $<$ Abulia on Transfer Items	.78
Abulia $<$ Tau Effect on Transfer Items	.89
Percent Correct	
Con. App. $<$ Tau Effect on Total Test	.78
Con. App. $<$ Abulia on Total Test	.67
Abulia $<$ Tau Effect on Total Test	.67

TABLE 9 (Continued)

Independent Variable			
Dependent Variable			
	Specific Relation	Proportion of Cases (N=9 unless indicated)	
	Con. App. < Tau Effect on Extension Items	.78
	Con. App. < Abulia on Extension Items	.67
	Abulia Tau Effect on Extension Items	.67
	Con. App. < Tau Effect on Transfer Items	.78
	Con. App. < Abulia on Transfer Items	.67
	Abulia < Tau Effect on Transfer Items	.89
Test Duration		
	Con. App. > Abulia	1.00
	Con. App. > Tau Effect	.78
	Tau Effect > Abulia	.89

Group analyses of percent correct performance were separated into three different dependent measures; total test score, extension score and transfer score. Each set of data were subjected to an arc sin transformation of proportions. The separate analyses revealed different relations between variables.

Both total test performance and transfer performance analyses did not reveal significant main effects of concept, order of study programs nor a significant interaction between these factors. However, the total test analysis indicated a significant main effect of study program. Analyses of study program effects were calculated by partitioning the order of study programs by concept interaction into study program and residual effects. An F test of this partition for total test performance yielded a significant effect of study program, $F(2, 12) = 3.95$, $p < .05$. Since there were neither order effects nor interaction effects, this ratio should be unbiased. As stated above, planned comparisons between the different study programs revealed a significantly lower performance after definition training than both example identification and exemplify training. However, there was no difference between example identification and exemplify training.

A similar ANOVA for correct percent transfer performance was also conducted. The ANOVA yielded a nonsignificant main effect of study program, $F(2, 12) = 2.19$, $p > .05$. All other effects and interactions were also nonsignificant.

Analyses of percent correct extension performance revealed different relations than the either total test or transfer analyses. The effect of order of study programs and the interaction between order and concepts were not significant. However, there was a significant effect of concept, $F(2, 12) = 4.15$, $p < .05$, and a significant effect of study program. $F(2, 12) = 4.53$, $p < .05$.

A 3 way ANOVA for total test duration was also conducted. This analysis revealed a significant effect of concepts, $F(2, 12) = 8.73$, $p < .01$. However, there was no significant effect of order of study programs and no interaction between order and concept. Partitioning the order-by-concept interaction into study program and residual effects did not reveal a significant effect of study program.

Group analyses of correct rate of responding were also separated into three dependent measures: total test scores, extension scores and transfer scores. Each set of analyses revealed different relations between variables.

Table 10 presents the source data for the ANOVA conducted for correct rate of responding on all test items. This analysis revealed a possible interaction between order and concepts as estimated by the residual effect, $F(2, 12) = 3.94$, $p < .05$. Other significant effects were found for concept $F(2, 12) = 21.05$, $p < .01$, and study program $F(2, 12) = 7.10$, $p < .01$. However, the test for these effects was positively biased (inflated) because the interaction effect contributes to the main effects, but not to the overall residual or error. Planned comparisons between concepts and study programs were described above. Briefly, performance on constructional approach was significantly lower than on either abulia or tau effect, and abulia was significantly lower than tau effect. In addition, total test scores after example identification training were significantly higher than after exemplify and definition training.

- - - - - - - - - - - - - -

SEE TABLE 10, PAGE 102

- - - - - - - - - - - - - -

Table 11 presents the source data for the ANOVA conducted for correct rate of performance on extension items. This analysis also revealed a possible interaction between order of study programs and concepts as estimated by the residual effect, $F\ (2, 12) = 4.50$, $p < .05$. Other significant effects were found for concept, $F\ (2, 12) = 7.5$, $p < .01$ and study program, $F\ (2, 12) = 6.72$, $p < .025$. However, again the test for these effects was inflated because of the possible interaction effect. The planned comparisons described above revealed a significant difference between extension performance on constructional approach and the other two concepts. In addition, a significant difference was found between performance after example identification training and after the other two programs.

TABLE 10

ANALYSIS OF VARIANCE FOR CORRECT RESPONSES PER MINUTE FOR THE TOTAL TEST

Source	Sum of Squares	DF	Mean Squares	F
Mean	109061.333	1	109061.333	358.54
Order (A)	1264.222	2	633.111	2.08
Error	1825.111	6	304.185	
Concept (B)	9723.555	2	4861.777	21.05***
AB				
Program (C)	3277.67	2	1638.830	7.10***
Residual	1820.55	2	910.27	3.94*
Error	2771.555	12	230.962	

* indicates < .05

** indicates < .025

*** indicates < .01

- - - - - - - - - - - - - - -

SEE TABLE 11, PAGE 104

- - - - - - - - - - - - - - -

Table 12 presents the source data for the ANOVA conducted for correct responses per minute on transfer items. This analysis revealed similar relations as those described for total test and extension performance. There were significant effects of concept and study program. In addition, there was no estimated interaction between order of study program and concept, these effects were unbiased. The planned comparisons between concepts indicated that constructional approach was significantly different from the other two concepts. The planned comparisons between study programs revealed a significant difference between performance after exemplify training and performance after the other two study programs.

- - - - - - - - - - - - - - -

SEE TABLE 12, PAGE 105

- - - - - - - - - - - - - - -

In summary, group analyses of percent correct performance, duration and rate of correct responding revealed a number of statistically significant relations. Table 13 presents each of these relations. A single asterisk indicates that the relation was significant at the .05 level, a double asterisk indicates that the relation was significant at the .025 level, and a triple asterisk at the .01 level.

TABLE 11

ANALYSIS OF VARIANCE FOR CORRECT RESPONSES PER MINUTE ON EXTENSION ITEMS

Source	Sum of Squares	DF	Mean Squares	F
Mean	179748.481	1	179748.481	116.07
Order (A)	15475.651	2	7736.92	5.00
Error	9292.000	6	1548.666	
Concepts (B)	20960.074	2	10480.037	7.15***
AB				
Program (C)	19671.63	2	9835.82	6.71**
Residual	13190.29	2	6595.14	4.50*
Error	17596.666	12	1466.388	

* indicates $<$ .05

** indicates $<$ .025

*** indicates $<$ .01

TABLE 12

ANALYSIS OF VARIANCE FOR CORRECT RESPONSES PER MINUTE ON TRANSFER ITEMS

Source	Sum of Squares	DF	Mean Squares	F
Mean	78840.037	1	78840.037	350.52
Order (A)	292.740	2	146.370	.65
Error	1349.555	6	224.925	
Concepts (B)	4794.962	2	2397.481	17.51***
AB				
Program (C)	2418.74	2	1209.37	8.83***
Residual	357.85	2	178.92	1.31
Error	1643.111	12	136.925	

*** indicates $<.01$

- - - - - - - - - - - - - - - -

SEE TABLE 13, PAGE 107

- - - - - - - - - - - - - - - -

Discussion

Introduction. The importance of this study would be grossly overstated if all the results reported in the previous section were discussed. Therefore, this discussion is restricted to the most salient features of the reported data and the study in general. First, those conclusions that can be made from the data are specified. Second, three methodological issues are described that limited the experimental conclusions. Finally, a summary of conclusions, changes in strategy and future directions in research suggested by the study are discussed.

Conclusions supported by data. Three conclusions can be unambiguously drawn from all the different measures of study and test performance. These were: a) the concept constructional approach was more difficult to learn than either abulia or tau effect, b) subjects generally performed better on types of tasks for which they received training (extension tasks) than those for which they had not received training (transfer tasks) and c) study programs in which subjects were asked to supply definitions did not facilitate test performance as well as either example identification programs or exemplify programs. Other, less clear, but equally important conclusions concerning the

TABLE 13

SUMMARY OF SIGNIFICANT RELATIONS BETWEEN STUDY PROGRAM AND DEPENDENT MEASURES OF TEST PERFORMANCE, AND CONCEPTS AND TEST PERFORMANCE

Independent Variable Dependent Measure	F	Planned Comparison	F
Study Program			
Rate of Correct Responses			
Total Test	7.10***	Ex. Id. > Exem. or Def.	12.77***
Extension Items	6.71**	Ex. Id. > Exem. or Def.	11.76***
Transfer Items	8.83***	Exem. > Ex. Id. or Def.	16.56***
Percent Correct			
Total Test	3.95*	Def. < Ex. Id. or Exem.	5.10*
Extension Items	4.53*	Exem. > Ex. Id. or Def.	5.97*
Transfer Items	-	-	-
Test Duration	-	-	-
Concept			
Rate of Correct Responses			
Total Test	21.05***	Con. App. < Ab. or T.E.	30.16***
Extension Items	7.15**	Con. App. < Ab. or T.E.	12.12***
Transfer Items	17.51***	Con. App. < Ab. or T.E.	29.37***
Percent Correct			
Total Test	-	-	-
Extension Items	4.15*	Con. App. < Ab. or T.E.	4.49*
Transfer Items	-	-	-
Test Duration	8.73***	Con. App. > Abulia or T.E.	12.35***

* indicates level of .05
** indicates level of .025
*** indicates level of .01

differential effects ofexample identification programs and exemplify programs are also discussed. The reader should refer to Tables 6, 8, 9 and 13 for summaries of the data that support these conclusions.

Relations between concepts. The most consistent general finding was the difference between constructional approach and the other two concepts. First, during study trials, subjects had lower rates of correct responding, lower percent correct scores and took longer to complete programs for the constructional approach. Second, during test trials subjects had lower rates of correct responding, lower percent correct scores and took longer to complete tests for the constructional approach. In addition, all of the test differences between constructional approach and other concepts were significant at the .01 level. Therefore, it was concluded that constructional approach was a more difficult concept than either abulia or tau effect.

This was obviously an important result for the purposes of this specific series of investigations. All data from the present study needed to be analyzed in light of this fact. Therefore, a number of the relations between study programs and subject performance became clearer when variability between subjects was partially atrributed to the difficulty of constructional approach. For example, exemplify study performance was errorless for all but two subjects; yet, both subjects received exemplify training with constructional approach. Further, all future research needs to consider the fact that constructional approach as programmed here was more difficult. Therefore, in order to investigate the effects of different study programs, either the existing concept differences need to be minimized or they should be systematically varied.

If it is not feasible to simplify the program for teaching constructional approach, it may need to be eliminated from subsequent studies. Alternatively, the concept could be retained as it is currently programmed to examine the interactions between different kinds of concepts and different kinds of study programs. This implies a change in the direction of future investigations and the use of different experimental designs. It would also require the development of materials for other concepts. Finally, it implies that it is possible to define different categories of concepts. The issues raised by this kind of investigation will be discussed in Chapter III. At present it was concluded that more control over concept differences was needed and it was best to revise the current materials for constructional approach.

This revision took the form of an integration of individualized and group techniques reported in Chapter II. These procedures were adopted from Markle (1967), Merrill and Tennyson (1974) and Chase (1980). The group comparisons between the three concepts included two different kinds of study programs: One that had only example identification questions and a second comprised of definition and exemplify questions. In addition, three kinds of dependent variables were measured: rate of correct responding, percent correct and duration. Since such procedures did not lead to materials that minimized concept differences, other options were considered. These options are discussed in Chapter III.

Relations between extension and transfer. The second general conclusion drawn from the data was that subjects performed better on extension items than on transfer items. Comparisons between the extension performance and transfer performance revealed that most subjects were more accurate and had more rapid rates of correct responding on extension items than on corresponding transfer scores. In addition, 15 of the percent correct extension scores were equal to or higher than 80% and 13 of the rate extnesion scores equalled or exceeded .80 responses per minute. However, only 4 out of 27 transfer scores were at or above these criteria. Therefore, in general it was concluded that subjects performed better on task types for which they had been trained than on task types for which they had not been trained.

This relation appears to be more powerful when the effect of concept is considered. The constructional approach was involved in five of the seven occasions on which percent correct transfer scores were higher than extension scores.

Four of the eleven rate transfer scores that were higher occurred with constructional approach. Six of the twelve percent correct extension scores that fell below 80%, and seven of the fourteen rate extension scores that fell below .80 responses per minute occurred with <u>constructional approach</u>. As described above, study performance on the constructional approach was low, indicating that subjects did not learn this concept as well as the other concepts. Therefore, it was not surprising that extension performance was also poor. The fact that transfer performance in some cases was higher than extension performance might be attributed to factors that were not controlled in this study. Moreover, there may be aspects of the <u>constructional</u> approach that affect the relation between extension and transfer differently than the other concepts. Nevertheless, when comparisons between extension and transfer were restricted to the concepts <u>abulia</u> and <u>tau effect</u>, there was a strong indication that extension performance was better than transfer performance.

The implications of these results are not as simplistic as they sound. For many of the concepts that students learn, teachers and curriculum designers train and test only on a single class of tasks (Semb and Spencer, 1976; Chase, Johnson and Keenan, 1977). It is assumed that these teachers expect students to generalize to other kinds of tasks. Obviously, some students can do this. Also, most students can generalize to some degree to new tasks. However, such generalization or transfer is not guaranteed. If high levels of performance on a particular task are required in order to progress through a curriculum, then the teacher should not assume that training on another type of task will be sufficient. However, this conclusion needs to be tempered until further research is conducted.

Comparisons of definition programs with other programs. The third data based conclusion was that definition programs did not facilitate test perforamnce as well as either example identification programs or exemplify programs. There were five dependent measures of test performance that were used to compare the effects of definition programs to the effects of both example identification and exemplify programs. These comparisons were: 1) rate of correct responding on all tasks of the test, 2) rate of correct responding on extension tasks, 3) percentage of correct responses on all tasks, 4) percentage of correct responses on extension tasks and 5) duration of the test. Performance on transfer tasks had been analyzed, but for reasons discussed below these analyses were not included in this conclusion.

Out of the five test comparisons between example identification programs and definition programs, three favored example identification training and two revealed no systematic differences between training. Therefore, it was concluded that example identification programs facilitated learning better than definition programs. Out of five test comparisons between performance after exemplify programs and after definition training two favored exemplify training, two revealed no systematic differences between training, and one favored definition training. Therefore, it is tentatively concluded that exemplify programs facilitated learning better than definition programs.

The conclusions that example identification programs facilitated learning better than definition training is consistent with previous findings. Watts and Anderson (1971) found that application tasks (similar to example identification tasks) facilitated test performance on naming tasks (Similar to definition tasks), recognition of example tasks and the identification of novel examples better than did naming tasks. Miller and Weaver (1976) added example and nonexample discriminations to a program of definition tasks and found increased achievement on novel example and nonexample discriminations. Reenan and Grant (1979) found that programs composed of definitions and example identification tasks faciliated performance better than definition tasks alone. Therefore, the trend indicated by this study is fairly conclusive.

The slight advantage found for exemplify over definition programs was novel. It does not appear that any formal comparison between these two types of tasks has been reported previously. While the data indicated that exemplify programs facilitated test performance better than definition programs,further investigations of exemplify programs are recommended. The major drawback to exemplify programs appeared to be the length of time it takes subjects to complete them. If this time constraint is sufficiently aversive to students it is not likely that they will complete these tasks. In such cases, the slight advantage of exemplify tasks will be eliminated. Therefore, the practical implications of these findings may well be limited. Further research needs to be completed to decrease the amount of time spent on exemplify tasks and to further compare exemplify with other programs.

Comparisons of example identification and exemplify programs. Having described the differences between concepts and the differences between definition and the other programs, the remaining critical comparisons are between example identification and exemplify programs. In addition, the differences between concepts are important here for it allows a more fine grained analysis of the differences between example identification and exemplify programs.

Comparisons of study performance revealed that example identification and exemplify programs affect different dependent measures differently. The rate of correct responding was higher and the duration of the study sequence shorter for most subjects after training in example identification. The one subject who responded at a slower rate on an example identification program did so with the <u>constructional approach</u>. Both exceptions to the comparison of study duration occurred with the <u>constructional approach</u>. However, subjects made less errors during exemplify programs and subsequently their percent correct scores were higher on exemplify programs. Therefore, subjects were able to respond faster to example identification tasks and to finish the example identification study program more rapidly, but they performed more accurately with exemplify programs.

It is difficult to conclude which of these positive effects is more important. The literature on errorless learning suggests that errors beget errors (Sidman and Stoddard, 1967) and therefore a program should avoid errors. However, as yet no experiments on errorless learning have been found by this authro on this kind of conceptual learning with this population. In addition, the correlation between numbers of errors and total test performance in this study was not significant, $r = -.15$, $p > .05$. Yet, both subjects who scored below 60% on the study program (Subjects 3 and 9) did poorly on the corresponding test (36% and 51% respectively). Moreover, it was statistically obvious that more subjects were needed before correlations between errors and test performance can be tested. Therefore, no definitive conclusions can be stated about the relative effects of errors.

A logical analysis might suggest that a time criterion is more important criterion than simple accuracy. Certainly there are many kinds of studying for which accuracy and speed are important. However, as long as the study behavior of interest is consequated differentially, and the number of errors is not so great as to become aversive to individuals (i.e., they stop studying), producing more correct responses per unit time might well be preferable. First, it allows for the response class to be practiced more often within a given time period. Second, it allows corrective feedback to be provided more frequently. Finally, it should shorten the amount of time that it takes to study; in other words it should be more efficient.

This conclusion must be tempered pending further investigations. Equally as important, there may be conceptual learning conditions which dictate that accuracy during study trials is critical. However, under conditions similar to those reported here in which subjects make a reasonable number of correct responses and are given feedback for both correct and incorrect responses, a time based criterion of study performance may be prefered to an accuracy criterion.

Comparisons of test performance after example identification and exemplify programs revealed that in general these two programs facilitated different kinds of performance. Most subjects' rates of correct responding were faster on extension tasks and on the total test after training on example identification. In addition, most subjects took longer to complete the test after exemplify training. However, exemplify training appeared to facilitate higher percent correct scores on extension items. On all other measures, comparisons between example identification and exemplify training did not reveal any systematic differences across subjects.

These test results, if reliably reproduced in other experiments, indicated that example identification and exemplify programs facilitate different kinds of performance. Therefore, no clear, single, general guidelines for preferred type of task to include in study programs can be offered. However, if the test results were analyzed in conjunction with study performance, it appeared that example identification tasks have a slight advantage over exemplify questions. The critical variable was time. Example identification programs were more efficient for the subject therefore might be the program of choice.

<u>Methodological issues</u>. Throughout the description of the results, a number of methodological problems were raised. In addition, other issues need to be addressed before the answers and directions that this study provided can be summarized. These issues have been organized into three separate categories. They are that: 1) the generally low extension performance suggests difficulties in interpreting transfer and total test comparisons, 2) the use of repeated

measures latin square experimental designs limits the interpretation of main effects if the're interactions between order and other main effects, and 3) the transfer performance comparisons that were analyzed were confounded by concurrent changes in tasks used to assess transfer.

The extension items were included in the transfer test in order to determine how well the subjects learned to respond to the kinds of questions for which they had received training. If they responded successfully, then it was appropriate to ask whether learning one kind of question facilitates performance on other kinds of tasks. However, what are the criteria for determining whether the trained task type has been adequately learned?

In the absence of any formal requirements, it is possible to select an arbitrary performance criterion. For percent correct a typical criterion is 90% correct (Block and Burns, 1976). If this criterion was used, then there were insufficient data to answer the transfer question. In fact, extension performance reached beyond 90% on only five of the twenty-seven tests and they all occurred with exexemplify tasks. Therefore, no transfer comparisons could be made. If the criterion were lowered to 80%, there were still too little data to conduct between-subject analyses and only one subject (Subject 1) with whom to conduct intrasubject analyses. Therefore, it appears impossible to answer the primary question for which this study was designed: does learning to respond to one kind of question facilitate performance on other kinds of questions.

At the very least, it is obvious that any conclusions about transfer performance would be tentative. In fact, because of the low extension performance, it is probably necessary to restate the primary question answered by this research. That is: what are the differential effects of exposure to three different study programs. The answers to this question will be discussed and summarized below.

The more important issue raised by this analysis is how to assure high enough performance on extension tasks in future studies so that the original, more important question can be answered. The following should accomplish this goal. First, further revisions can be made on the study programs to make it more likely that subjects will learn the tasks included in the program. Second, a mastery criterion should be implemented for study performance. If subjects are required to perform correctly on a broader range of questions in each study program, then they are more likely to respond correctly on extension items (Morrisett and Howland, 1959). Third, motivation to respond correctly could be increased by implementing reinforcement contingencies during both study and test trials. One reason subjects may have performed poorly was that the arbitrary conditions of the laboratory setting and the concepts used interferred with the usual reinforcers gained from learning verbal responses. Therefore, arbitrary consequences like payment for correct responses may be necessary to increase and maintain the subjects' behavior. These procedures should make it more likely that the original questions about transfer can be answered in future research.

The second methodological issue that needs to be addressed is the possible problems inherent in repeated measures, latin square designs. A repeated measures design was used because it is one way that intrasubject analyses can be combined with group analyses to obtain as much critical information as possible from the smallest number of subjects. It is important, however, to assume that carryover effects or effects due to position in time are minimal. Carryover effects were expected to be small or nonexistent because of the differences between concepts, the differences between conditions and because the intersession intervals were long enough to anticipate substantial forgetting. In addition, previous studies have failed to show a significant effect of position in time (Chase, 1980). However, it was impossible to separate the carryover effects from the concept effects in this study. The order of concepts was not varied because of the difficulty of interpreting data when two order variables are manipulated. Order of study programs was varied instead. Therefore, since the constructional approach always occurred on the last session, and since performance on the constructional approach was lower than the other concepts, performance on the third session was also lower. Future studies need to separate the effects of position in time from the effects of concepts.

A more critical limitation can be attributed to the latin square aspect of the design. There are two major problems with latin square designs: 1) that main effects and interaction effects are confounded and 2) that if an interaction between an order or sequence variable and any other factor is revealed, the F ratios for the other factors will be biased. In this study, the F ratios for the effects of study program and concept would be positively biased (inflated) if there was an interaction between order of study program and concepts. However, this interaction can only be estimated because the interaction variability is at least partially due to the effect of study program. An estimate is obtained by subtracting the study program variability from the order by concepts interaction and submitting the resulting residual variability (B cell residual) to an F test. If the test is nonsignificant, it is assumed that variability due to the interaction is negligable.

Unfortunately, two of the analyses in this study revealed a significant F ratio for the B cell res. This resulted in F ratios that were inflated for both rate of correct performance on total test and rate of correct performance on extension items. However, both B cell res. F ratios were barely significant at the .05 level (i.e., $F(2, 12) = 3.90$ and 4.47 respectively) and the F ratios for the main effects of study program and concept for each of these measures were high (i.e., F (2, 12 = 7.10 and 6.71 respectively for study programs and 21.05 and 7.15 for concepts). Therefore, though it is impossible to draw completely unambiguous conclusions for these two measures this author concludes that the main effects of concept and study program did exist.

A third methodological problem was due to the transfer measures used in this study. In general, transfer was defined as performance on any task type that was not explicitly taught during the study program. Thus, example identification tasks, exemplify tasks and combination tasks were considered transfer items after definition training and exemplify tasks. Definition tasks and combination tasks were considered transfer items after example identification training. The subject's performances on each of these tasks were combined to yield one transfer score for each test. It should be clear, then, that the effects of the independent variables are confounded by the effect of the task types being measured. The kinds of tasks included in a measure of transfer changes systematically when training changes. If differences are found, it could be because of the differences in training, but it may also be due to differences in the test tasks themselves. If no differences in task types were found on comparisons of study performance and extension or if some kind of quantitative relation could be specified (e.g., responses to example identification tasks were five times faster than exemplify tasks), then the differences between transfer performance could be analyzed in terms of study program effects. However, as neither of these conclusions can be supported, the measure of transfer described have can not be used as a means of comparing study programs.

This problem can be eliminated by making sure that the tests include at least one class of tasks that is not a component of any of the study programs compared. In this case, the combination task would be used as a critical test of transfer because it is a transfer item for all three conditions. These analyses were conducted in Experiment 2. Although an analysis of combination performance could be extracted from the data collected here, the author decided that little would be gained by this effort. First, the problem of interpreting transfer performance given low extension performance would still exist. Second, the conclusion that the overall transfer data can not be compared was made after the data from experiment 2 were analyzed. As shown in Experiment 2, performance on combination tasks was low and did not vary as a function of study program.

Therefore, the combination data for this study were not separated from the other transfer data and analyzed separately.

In summary, three methodological problems which may limit some of the conclusions that can be made were discussed. The first problem, that extension performance was low and therefore did not indicate mastery of the trained task types, prevented answering the original question of the study. However, revisions in study materials, a mastery criterion during study trials and motivational contingencies for correct performance should minimize this problem in future research. The second problem, that any interactions would bias the results in a latin square design, appeared to affect only two analyses, rate of correct extension performance and rate of correct responding on the total test. Even in these cases there was sufficient evidence that the interaction effect was small enough and the main effects large enough to conclude that the main effects did, in fact, exist. The third problem, that comparisons of transfer were confounded by concurrent changes in the tasks that define transfer, prevented the use of transfer performance as a means of comparing study programs. It was impossible to determine whether the differences found on transfer performance were due to the type of training or the type of transfer tasks. One possible solution to this problem is to analyze performance on each type of task separately and to include at least one class of tasks that is a transfer task regardless of training. This can be accomplished in future studies by analyzing performance on combination tasks separately.

Thus, the limitations of this study have been specified. Conclusions based on rate of correct extension performance and rate of correct performance on the total test must be drawn with caution because of the interactions found. In addition, transfer differences can not be used as a criteria for comparing study programs or concepts.

<u>A summary of conclusions and future directions</u>. Throughout this discussion a variety of tentative conclusions have been posited. To summarize: first, the concept constructional approach was concluded to be more difficult than the other concepts used in this study. Thus, the concept was revised for subsequent use. Second, extension performance was found to be higher than transfer performance. Thus, teachers should not assume that training on one class of task will faciliate transfer to other classes of tasks. Third, example identification programs appeared to be superior to definition and exemplify programs.

However, these conclusions were stated tentatively because of the three methodological problems that have been discussed. First, transfer performance could not be included as a factor in these conclusions because of the possible task confounding variable that was present. Therefore, there was one less, critical criteria for deciding between types of study programs. Second, the possible interactions between concept and order of study program prohibited conclusive analyses of the main effects of the three study programs on some measures. Most importantly, the low extension performance allowed for the possibility that if subjects had learned each of the tasks better, that relative test performance would have been different. Attempts to eliminate these problems in future studies will make future results less ambiguous.

In conclusion, future investigations should include a replication that is revised in terms of methodological changes recommended above. Also, new ways of improving example identification performance and comparisons of other kinds of study programs with example identification programs should be pursued. These suggestions indicate three different directions for future studies. The third option was pursued and reported below (Experiment 2) for a variety of reasons. First, a number of previous studies have demonstrated the effectiveness of example identification type questions. Second, it was not evident that example identification questions should be pursued exclusively. Example identification programs may faciliate learning but they were time consuming to design, test and implement. Other kinds of questions were easier for teachers to write, therefore it may be beneficial to look at how other tasks can be combined in a program. Third, the effectiveness of the example identification programs can be improved upon and the methodological changes can be made while looking at other relations. Therefore, the next study compared example identification programs to programs comprised of multiple question types. Specifically, example identification programs were compared to two programs: 1) a combination of definition and example identification tasks and 2) a combination of definition and exemplify questions.

CHAPTER III

EXPERIMENT 2

Purpose

The purpose of this second experimental test of transfer of learning was to compare the example identification programs to programs comprised of multiple question types. Two programs were selected: one with definition and exemplify questions and one with definition and example identification questions.

The example identification program was selected because previous studies, including the first study reported here, have indicated that example identification tasks or similar tasks were better for most dependent variables than other single task types. However, there were problems with examples identification programs that indicate the need to compare them to new kinds of programs. Example identification tasks were exceedingly time consuming to write and always required substantial testing and revision. In addition, test performance could still be vastly improved. Therefore, it is important to continue investigating example identification questions.

The definition/exemplify program was selected for two reasons: 1) investigations of a combination of such questions were not found in the literature and 2) the possibility that combining these two tasks would eliminate the slight advantages of example identification programs. Specifically, the rate of answering exemplify questions might be increased if preceded by definition questions that assured that subjects have learned the critical features of the concept. Conversely, answering exemplify questions might extend the subjects repertoire from being able to state the general rule (i.e., definitions) to responding to other types of tasks. In addition, a new type of exemplify question asked the subjects to list the irrelevant properties of the concept before requiring them to integrate the irrelevant properties with the general rule learned from the definition tasks. For example, the irrelevant properties of abulia were the specific behaviors, people or animals, and reinforcers for these behaviors and individuals. Therefore, subjects were asked to list six behaviors, six animals and six possible reinforcers before they were asked to use these features in examples of abulia.

The second multiple task type program, the define/example identification program, was elected for reasons similar to the define/exemplify program. A number of studies have compared define-type questions to example identification-type questions (cf., Watts and Anderson, 1971; Miller and Weaver, 1976; Keenan and Grant, 1979). However, few of the studies reviewed looked at the effects of combining tasks in a program. Miller and Weaver (1976) looked at a combination of define and example identification tasks, but this condition was compared only to definitional questions alone. In addition, the results of this study may have been confounded by sequence effects. All subjects received the various conditions in the same order. Therefore, it was impossible to analyze the contribution made by order of treatment. Keenan and Grant (1979) also looked at combinations of define and example identification questions. They did compare this condition to example identification questions alone. However, they did not compare define/example identification programs to other multiple task type programs. In addition, Keenan and Grant (1979) did not control for differences in the difficulty of the tasks on the probe tests. Therefore, the slight differences that were found between conditions could have been due to some tests being more difficult than others.

Thus, this study was the next logical step. Study programs or task combinations that have been studied only under incomplete methdological conditions were compared to the type of task that has proven both in the previous experiment and throughout the prose learning literature to be the most effective question type. These comparisons were made within an experimental design that controls for many of the other variables that may influence the relations between study programs and test performance. Controls for motivation, concept difficulty, task difficulty and sequence effects were implemented. In addition, transfer performance was measured and analyzed for combination tasks separately in order to eliminate the task confounding variable discussed in Chapter II. Other variables such as pre-experimental study skills and entering behavior with respect to the concepts and features of the concepts used in the experiment were monitored. It was hoped that these changes would permit more definitive conclusions to be drawn from this study.

Methods

Subjects. Eighteen undergraduates recruited from a special introductory psychology course for transfer students at the University of Massachusetts served as subjects. None had any previous experience with the experimental procedures or concepts. All were sophomores and juniors in college, majoring in psychology, and had mastered introductory level concepts in basic learning and principles and experimental methdologies before the experiment took place.

Personnel. The graduate student who conducted the first experiment also coordinated experiment 2. Four undergraduate psychology majors served as research assistants. All four conducted experimental sessions, corrected the transfer tests and helped compute the data. Two assistants helped to analyze the data, transcribed the tapes, checked the reliability of implementing the experimental procedures and rescored the transfer tests for calculating indices of agreement. Training for all assistants was identical to that in experiments 1.

Setting. The setting was the same as that used for experiment 1.

Materials. Materials were similar to those used in experiment 1. However, the materials for constructional approach were revised extensively. Five subjects were exposed to all the tasks and prose passages written for the three concepts. The experimenter presented the task, asked for an answer, corrected the answer and, finally, discussed the task and answer with the subject. The subject was asked to critique the task. Specific questions were asked about each task (e.g., "Do you think the task was too long?" "Was it amusing?" How confident were you that you were correct?"). This information was used as a basis for rewriting the tasks. Subjects were also asked to compare the three concepts. After three sets of revisions, the responses to all those concepts were similar, thus permitting within-concept task difficulty to be assessed. However, the critical test of concept similarity requires between group comparisons with more subjects. Experiment 2 constituted this group comparison.

Three additional material changes were made. First, the test for each concept was standardized with an equal number of example identification tasks on each test. (See Table 14). Second, subjects were no longer required to estimate the amount of time spent on different kinds of study behavior for the Study Behavior Questionnaire. Instead, a simple "yes" or "no" response was requested by the questions. Appendix H illustrates the new Study Behavior Questionnaire.

Third, questions on the simpler concepts that were critical features of the concepts abulia, tau effect and constructional approach were included on the pre-test. These simpler concepts had been presented in the introductory psychology course from which the subjects were recruited. Previously, it had been assumed that subjects were able to answer questions related to these concepts. However, during experiment 2 this assumption was explicitly tested. Appendix I illustrates the new pretest.

- - - - - - - - - - - - - - -

SEE TABLE 14, PAGE 134

- - - - - - - - - - - - - - -

All other experimental materials and apparatus were the same as for Experiment 1.

Procedures. Each subject participated in four, one hour sessions. Session one was similar to the first session of Experiment 1 with two exceptions. First, the items on the Study Behavior Questionnaire were answered after subjects had finished studying the 900 word passage. Second, subjects were told that they would be paid ten cents for each correct response on the pre-test and all subsequent tests. The second through fourth sessions were also similar to experiment 1. Again, the differences between experiments 1 and 2 were that subjects were paid ten cents immediately after each correct study answer and at the end of the experiment they received a check for the total earned on the four tests.

TABLE 14

DESCRIPTION OF PROBE TESTS FOR EACH CONCEPT REGARDLESS OF STUDY CONDITION FOR EXPERIMENT

Constructional approach	Abulia	Tau effect
A. 18 examples* and non-examples: - 9 of each	A. 18 examples and non-examples; 9 of each	A. 18 examples and non-examples; 9 of each
B. 8 define tasks; each feature rquired twice	B. 8 define tasks; each feature required twice	B. 8 define tasks; each feature required twice
C. 8 exemplify tasks; each feature required twice	C. 8 exemplify tasks; each feature required twice	C. 8 exemplify tasks; each feature required twice
D. 2 combination tasks; each with 4 parts	D. 2 combination tasks; each with 4 parts	D. 2 combination tasks; each with four parts

* All examples and nonexamples selected from the most difficult items from previous item analysis

Each subsequent session (2-4) was different only with respect to the type of study program that was manipulated. These differences are specified in Table 15.

- - - - - - - - - - - - -

SEE TABLE 15, PAGE 136

- - - - - - - - - - - - -

<u>Experimental design</u>. Two changes in experimental design were made from experiment 1. First, nine of the eighteen subjects were not given a pre-test. Second, the order of concepts was changed to separate the effects due to different concepts from the effects due to position in time. Therefore, the subjects were assigned to groups that varied according to whether or not they received the pre-test, and according to the order of training, and the order of concepts was different from the order used in experiment 1. These groups were counterbalanced. Table 16 illustrates this design.

- - - - - - - - - - - - - -

SEE TABLE 16, PAGE 137

- - - - - - - - - - - - - -

<u>Reliability and interscorer agreement</u>. Interscorer agreement procedures were identical to those used in experiment 1. In order to determine whether the experimental procedures were reliably implemented, two procedures were used. First, approximately 45% of the pretest sessions and 22% of the study

TABLE 15

DESCRIPTION OF THE THREE EXPERIMENTAL CONDITIONS: EXAMPLE IDENTIFICATION PROGRAM, DEFINE/EXEMPLIFY PROGRAM, AND DEFINE/EXAMPLE IDENTIFICATION PROGRAM

Example Identification	Define/Exemplify	Define/Example Identification
1. Prose passage	1. prose passage	1. Prose passage
2. copy tasks	2. copy tasks	2. copy tasks
3. example	3. define 1 (feature 1)	3. define 1
4. matched* non-example 1	4. define 2 (feature 2)	4. define 2
5. nonexample 2	5. define 3 (feature 3)	5. define 3
6. matched example	6. define 4 (feature 4)	6. define 4
7. nonexample 3	7. define 5 (feature 5)	7. define 5
8. example	8. exemplify 1	8. example
9. matched non-example 4	9. exemplify 2	9. matched non-example 1
10. random sequence of example and non-examples	10. exemplify 3	10. nonexample 2
.	11. exemplify 4	11. matched example
.	12. exemplify 5 (all features)	12. example
.	13. define 5	13. matched non-example 3
.	14. terminal	14. matched non-example 4
15.	15. terminal define	15. terminal define
16. Test	16. test	16. test

(During study sequence, whenever S's response was incorrect, E gave another tasks of the same type and number. However, the new task was given later in the sequence.)

* a matched illustration was identical to a preceeding illustration except for one critical feature.

TABLE 16

SEQUENCE OF EXPERIMENTAL CONDITIONS FOR EACH SUBJECT

Subject	Session 1 Construction-al Approach	Session 2 Abulia	Session 3 tau effect
	Pretest		
1	Example Id	Def./Exem.	Def./Ex Id.
2	Ex Id	Def./Exem.	Def./Ex Id.
3	Ex Id	Def./Exem.	Def./Ex Id.
4	Def./Ex Id	Ex Id	Def./Exem.
5	Def./Ex Id	Ex Id	Def./Exem.
6	Def./Ex Id	Ex Id	Def./Exem.
7	Def./Exem.	Def./Ex Id	Ex Id.
8	Def./Exem.	Def./Ex Id	Ex Id.
9	Def./Exem.	Def./Ex Id	Ex Id.
	No Pretest		
10	Ex Id	Def./Exem.	Def./Ex Id.
11	Ex Id	Def./Exem.	Def./Ex Id
12	Ex Id	Def./Exem.	Def./Ex Id
13	Def./Ex Id	Ex Id.	Def./Exem.
14	Def./Ex Id	Ex Id.	Def./Exem.
15	Def./Ex Id	Ex Id.	Def./Exem.
16	Def./Exem.	Def./Ex Id	Ex Id.
17	Def./Exem.	Def./Ex Id	Ex Id.
18	Def./Exem.	Def./Ex Id	Ex Id.

program sessions were examined to determine the assistants compliance with the prescribed procedure. Second, approximately 30% of the subjects' study answers were rescored. The agreement index for the pretest session was determined by comparing how procedures were actually implemented with those relevant items outlined in Table 2. (items II C., D., G., H., K., L., M., N., and R). The experimenter listened to the tapes for each session and when the assistant engaged in one of these behaviors a "+" was scored. The number of pluses was divided by the total number of items and multiplied by 100 to obtain percent agreement. The median agreement for the eight rescored pretest sessions was 88 with a range of 80-100% agreement the mean of 88.7%.

The agreement index for the other sessions was more difficult to assess as the assistants were not required to talk as much as in the pretest session. However, three criteria were applied. First, the assistants were to state explicitly that the subjects would receive ten cents for each correct response on both the study questions and the test questions. Second, corrective feedback was to be given after every study question. Third, no additional information was to be given about the concepts being learned. For instance, if additional examples or analogies were given or if the definition was paraphrased, the session was not implemented as prescribed. Two subjects were not told about the monetary incentive for correct study answers. However, in both cases they did receive the consequences for correct responding and were told that they would receive ten cents per correct response on the test. There were no instances in which corrective feedback was withheld. However, there were substantial differences in the style in which feedback was given. One assistant simply read the answers from the answer key, whereas the other assistants stated what was on the answer key and then elaborated or put the answer in their own words. In four out of fifty-four cases, these elaborations included additional information. It was impossible to determine whether these differences in procedure had any effects on the results. The answers to study questions were rescored in order to determine whether the assistants had correctly scored the answers. The interscorer agreement indices were calculated for each study session by dividing the number of agreements by the total number of items scored and multiplying by 100. The median agreement was 88%, the mean agreement was 87.82%, and indices ranged from 55-100%. The one session which received the score of 55% agreement was rescored a third time by the experimenter. The agreement between the experimenter and the second scorer was 88%. Therefore, the score given by the second scorer was used in subsequent analyses.

Interscorer agreement indices were calculated for each test measure as well. The same scoring procedure described in experiment 1 was used. The agreement indices were calculated by rescoring approximately 20% of the tests. Median agreement on example identification items was 100% ($\overline{X}$ = 98%); median agreement on definition items was 87% ($\overline{X}$ = 83.96%); median agreement on exemplify items was 100% ($\overline{X}$ = 94.58%) and the median agreement on combination tasks was 80% ($\overline{X}$ = 67.92%). The range for all these items was between 50-100%.

The poor agreement on both definition and combination tasks required that these tasks on all tests be rescored by the experimenter. A second agreement index was then calculated for these two task types by comparing the experimenters score to the second scorers. The mean agreement for definition tasks was 91.66%. The mean agreement for combination tasks was 86.08%. The experimenter's scores were then used in all subsequent analyses.

In addition to these measures of interscorer agreement, an index of interobserver or interrecorder agreement was calculated for the recording of durations. A second observer observed 5 of the 54 experimental sessions through the one-way mirror connecting the two carrels. The second observer activated the timer when the subject was engaged in each task and recorded the time it took the subject to finish the task. These times were compared to those recorded by the research assistant. Durations were considered to be in agreement if they were within $\pm$ 2 seconds of each other. The median agreement was 82% with a range from 70-100% and a mean of 84.20%.

Results

The results were analyzed in terms of four independent variables: exposure to the pretest, study program, concepts and the order in which study programs were presented. Data were separated into study and test performance. Analysis of study and test performance were further divided into the measures of rate of correct responding, percent correct and duration. Rate and percent correct test performance were reported in terms of total test performance, extension and transfer performance (i.e., combination tasks). These components of the test were analyzed separately for three reasons. First, total test performance was the most meaningful measure because the combined performance on all four task types could be directly compared and because it was most analogous to a classroom evaluation. Second, extension performance was examined to determine whether the specific learning acquired during the study sequence was maintained on the test and to compare extension to transfer performance. Third, transfer performance was isolated to determine if the learning acquired during the study program facilitated performance on a task type not included in that program. Duration measures were reported only for total test. In addition, inter and intra-subject comparisons were conducted for both study and test performance analyses. Other factors and interactions that possibly effected the exceptions to the intrasubject relations were analyzed as well.

Study performance. Figures 10-15 present the data for correct responses per minute for all eighteen subjects. Individual graphs group subjects according to the combination of study program and concept that they received. Figures 10-12 present the rate data for those subjects who received a pretest. Figures 13-15 present the rate data for those subjects who did not receive a pretest. The black circles represent correct responses and the open circles represent incorrect responses. The time in minutes indicates the time subjects spent answering the questions (duration). Duration does not include the time required to present the questions nor the time required to give the subjects feedback on their answers. The y axis depicts the number cumulative of correct responses.

- - - - - - - - - - - - - - - - - -

SEE FIGURES 10-15, PAGES 143-160

- - - - - - - - - - - - - - - - - -

The rate data demonstrated several systematic relations. Differences were found among the study programs while no differences were found to be related to exposure to the pretest, order of presenting the study programs or concepts.

First, Figure 10-15 reveal that the rate of correct

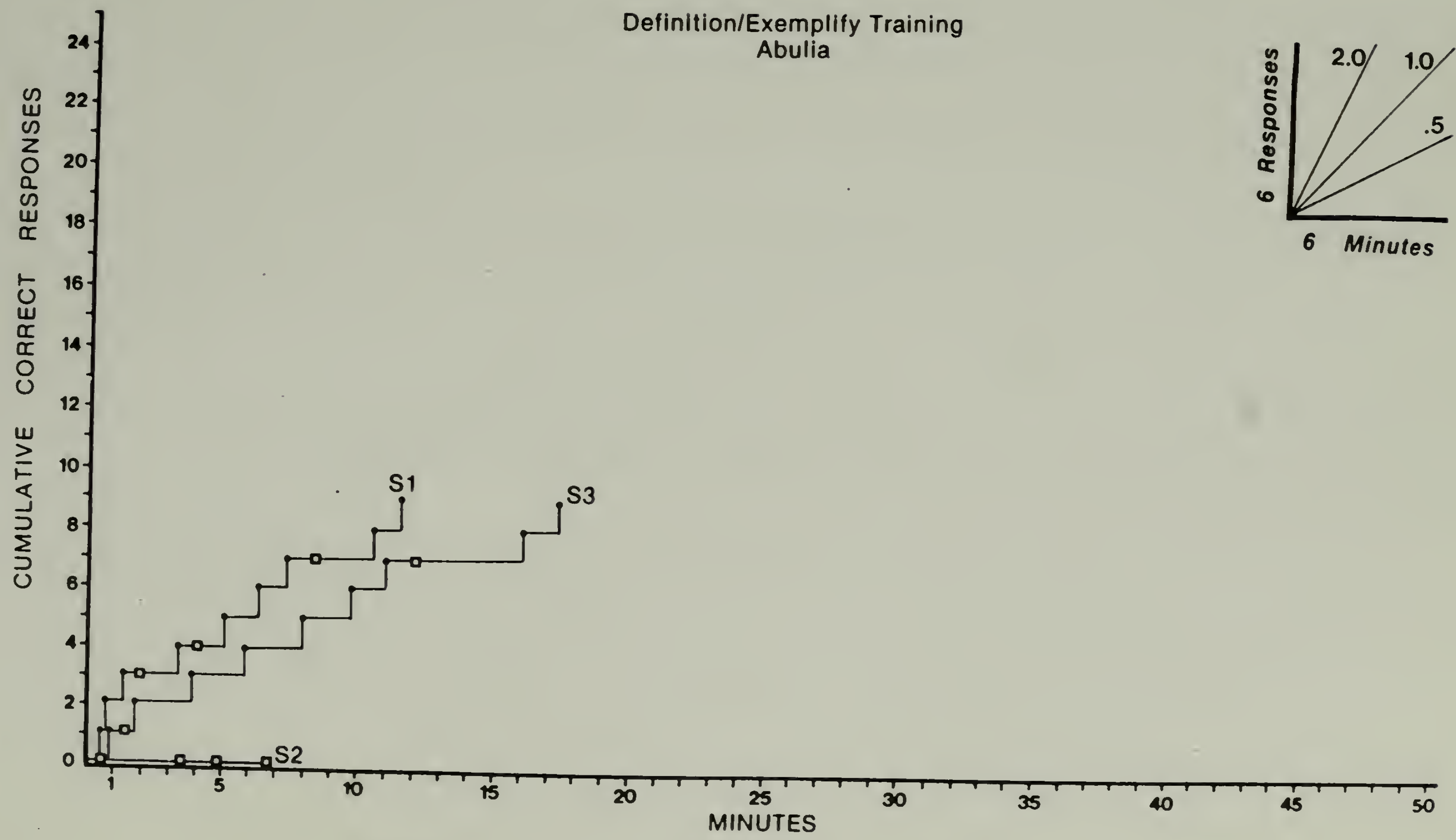
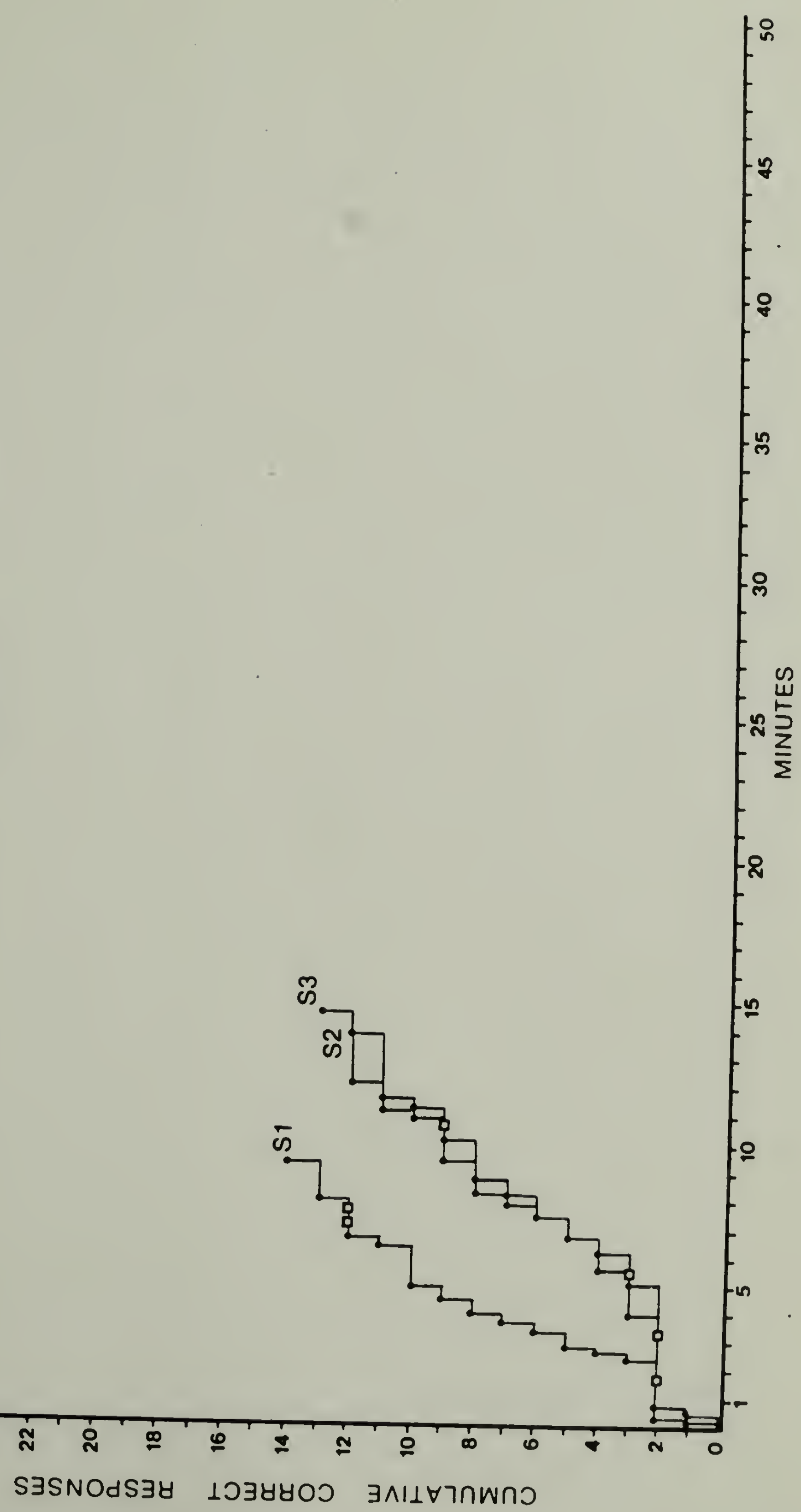
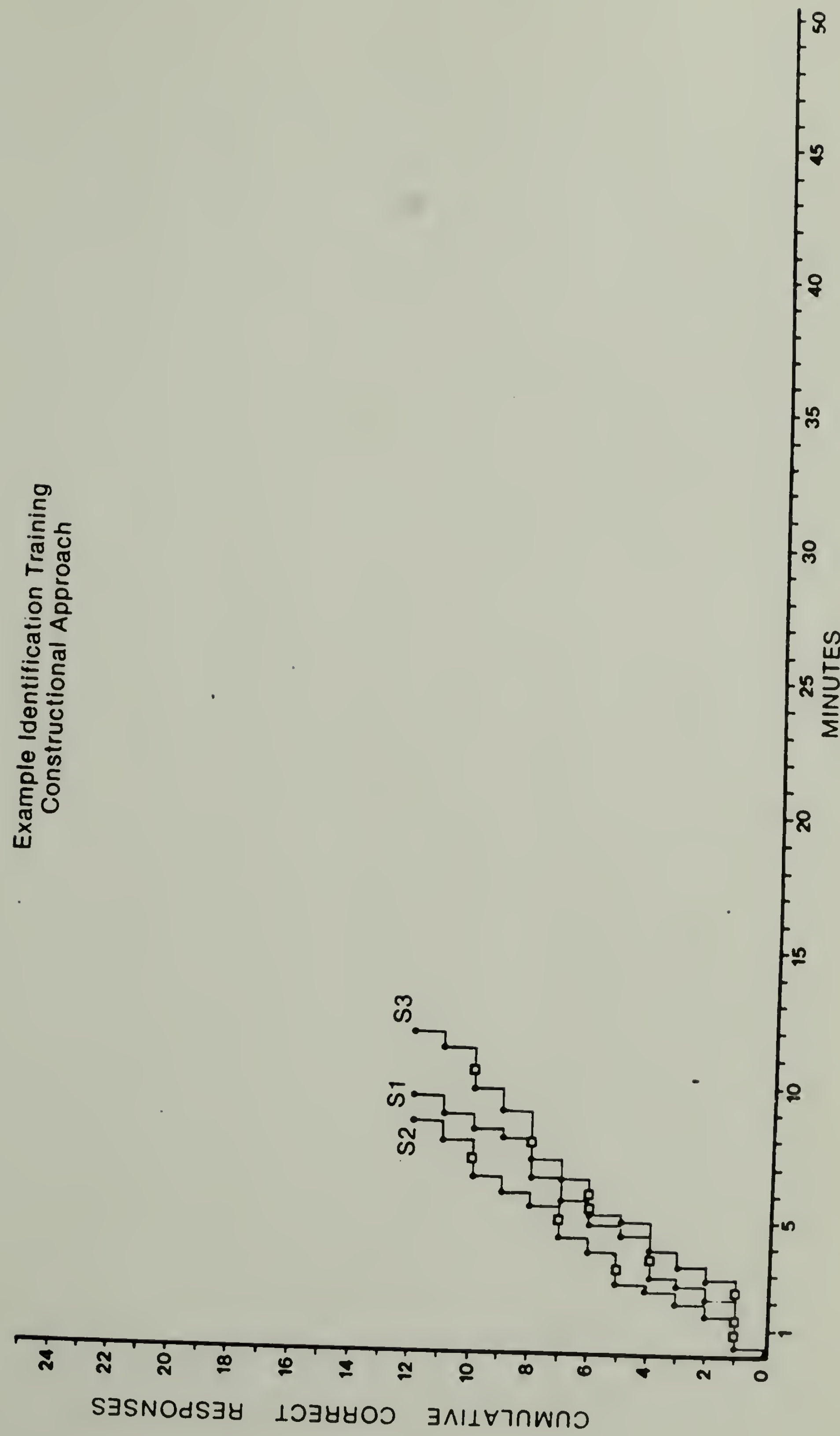


Figure 10. Cumulative frequency of study performance.

Definition/Example Identification Training
Tau Effect





Example Identification Training
Constructional Approach
CUMULATIVE CORRECT RESPONSES
MINUTES
S1
S2
S3
0
2
4
6
8
10
12
14
16
18
20
22
24
1
5
10
15
20
25
30
35
40
45
50

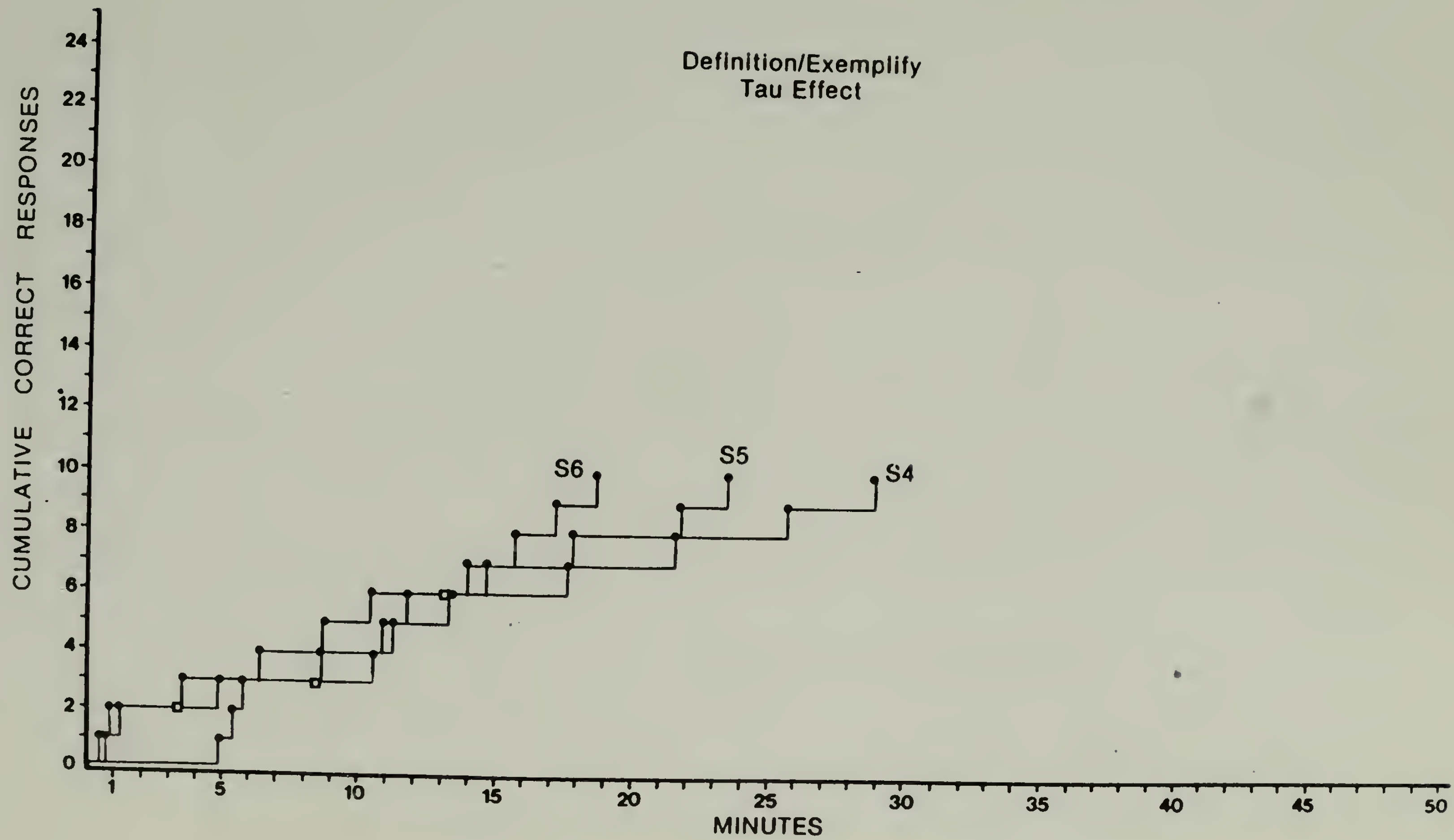
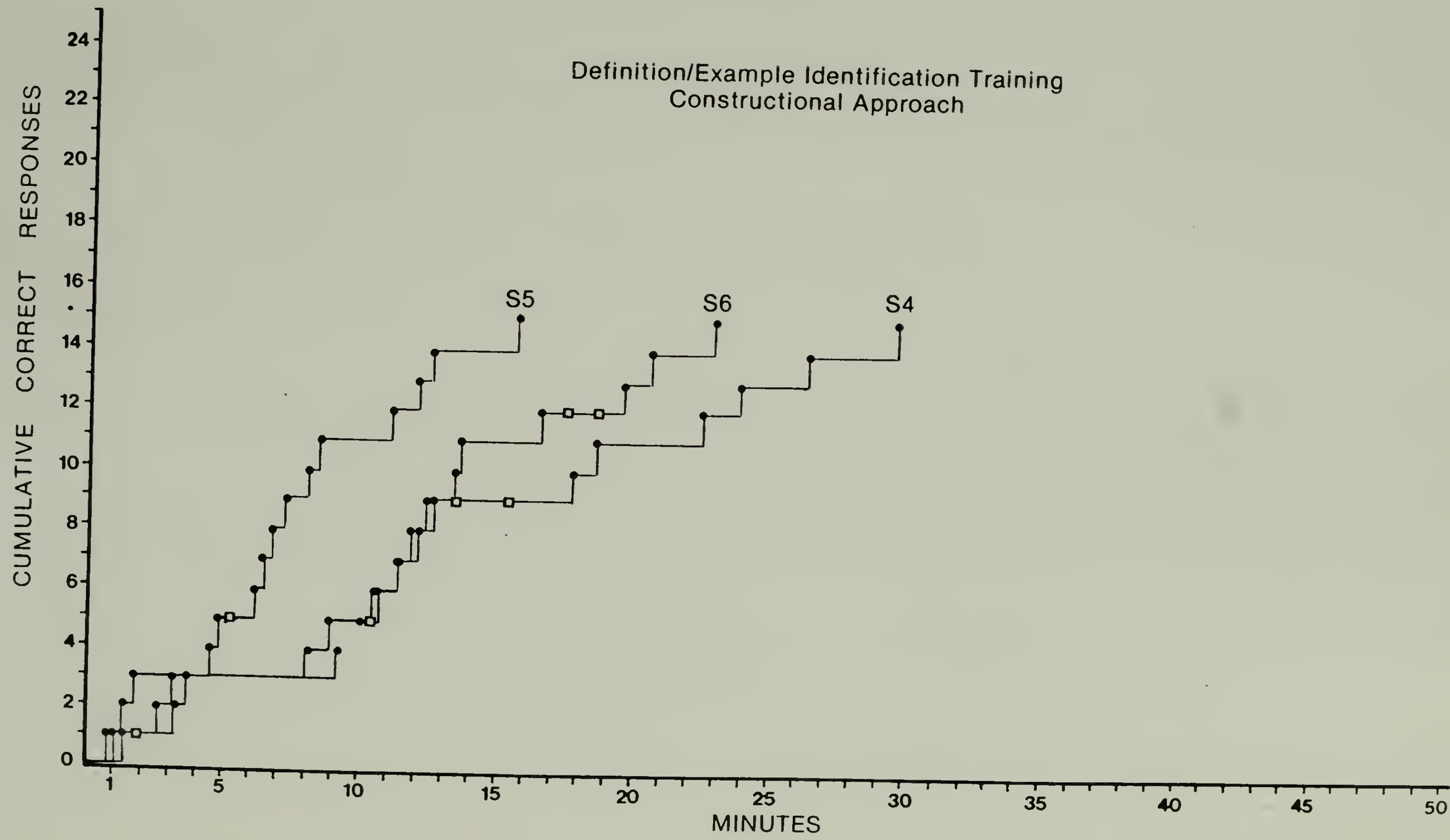
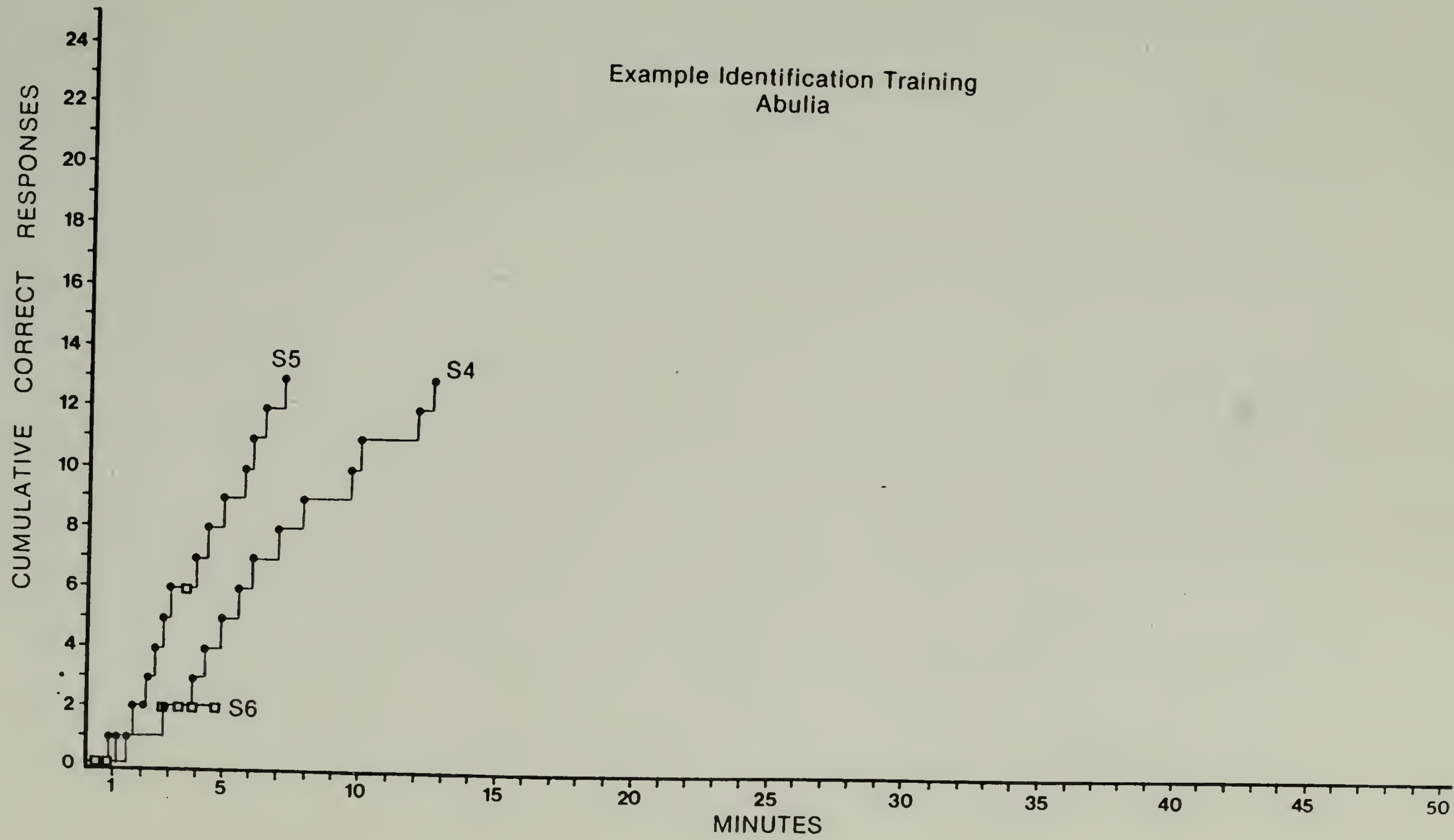


Figure 11. Cumulative frequency of correct study performance.



Definition/Example Identification Training
Constructional Approach
S5
S6
S4
CUMULATIVE CORRECT RESPONSES
0
2
4
6
8
10
12
14
16
18
20
22
24
1
5
10
15
20
25
30
35
40
45
50
MINUTES


Example Identification Training
Abulia
CUMULATIVE CORRECT RESPONSES
0
2
4
6
8
10
12
14
16
18
20
22
24
MINUTES
1
5
10
15
20
25
30
35
40
45
50
S5
S4
S6

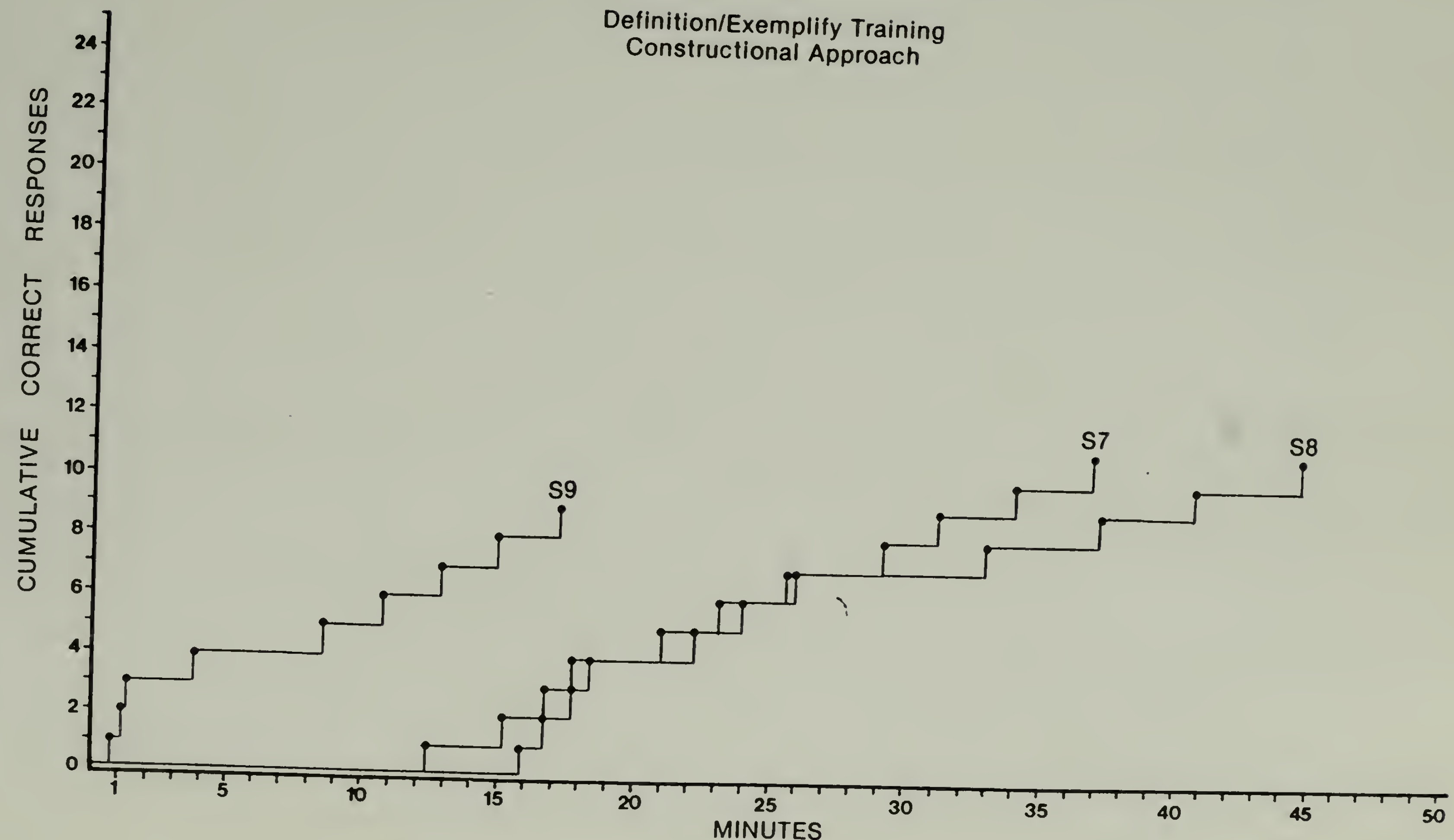
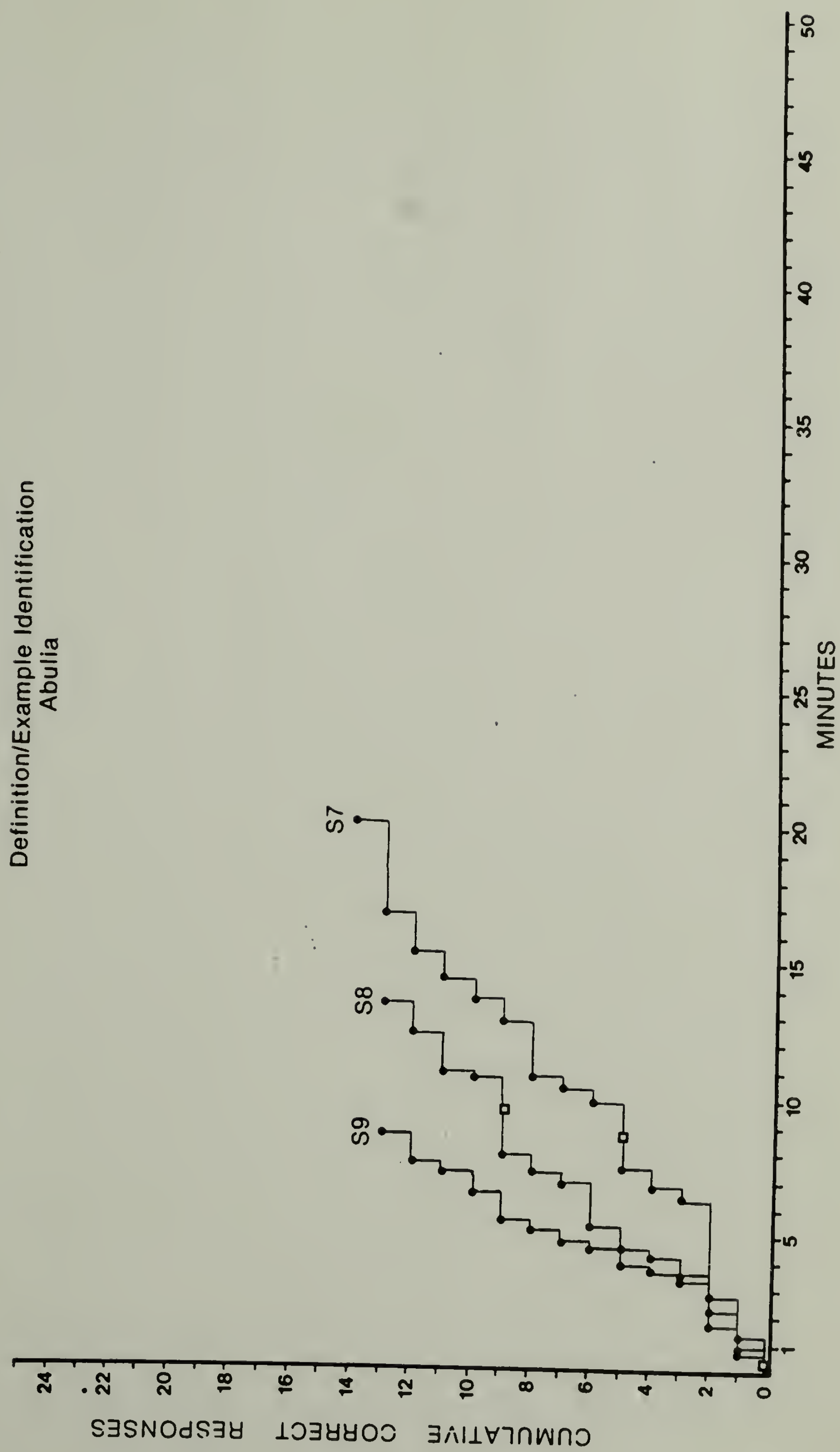
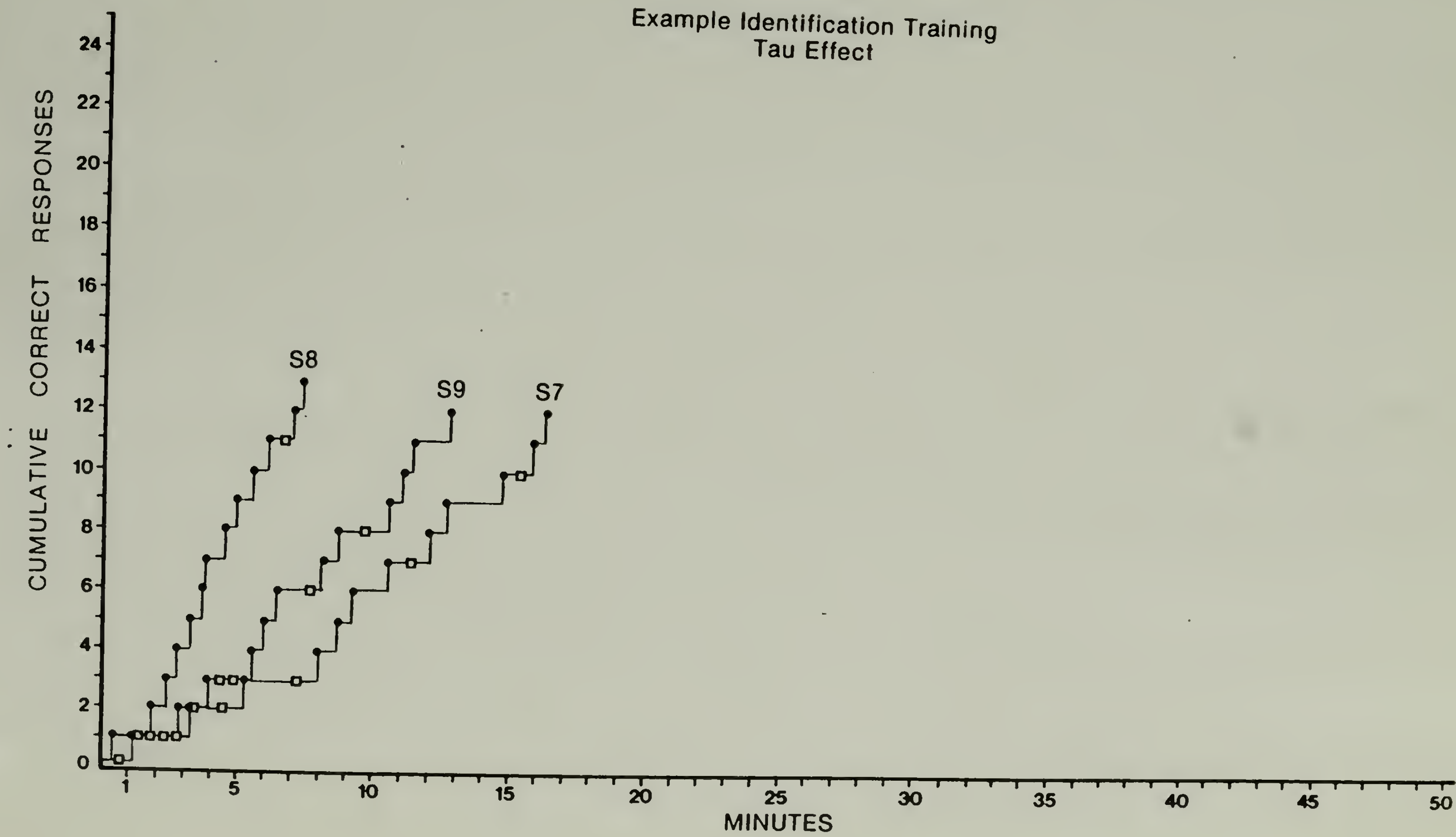


Figure 12. Cumulative frequency of correct study performance.


Definition/Example Identification
Abulia
CUMULATIVE CORRECT RESPONSES
MINUTES
S7
S8
S9
0
2
4
6
8
10
12
14
16
18
20
22
24
1
5
10
15
20
25
30
35
40
45
50


Example Identification Training
Tau Effect
CUMULATIVE CORRECT RESPONSES
0
2
4
6
8
10
12
14
16
18
20
22
24
MINUTES
1
5
10
15
20
25
30
35
40
45
50
S8
S9
S7

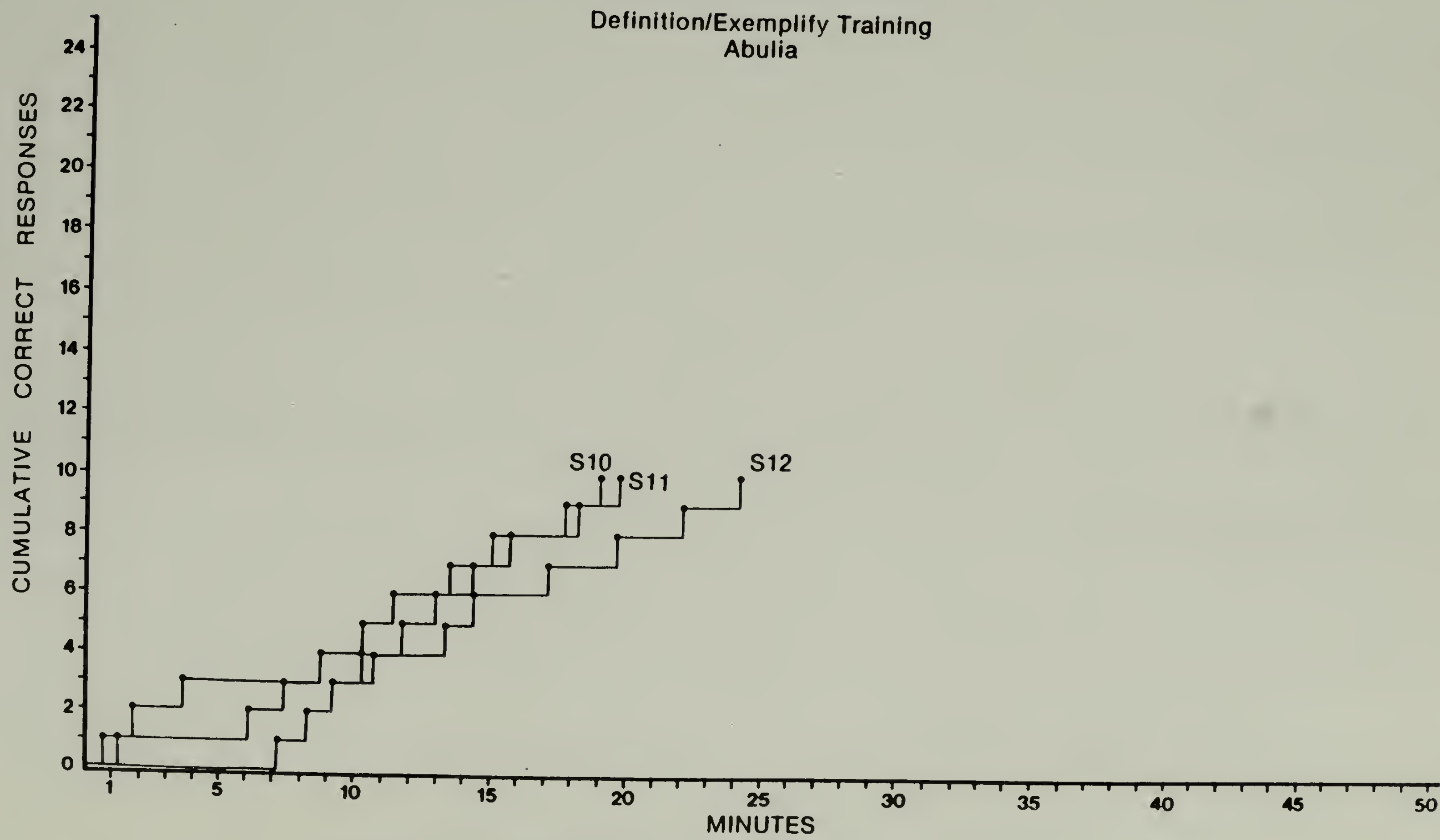
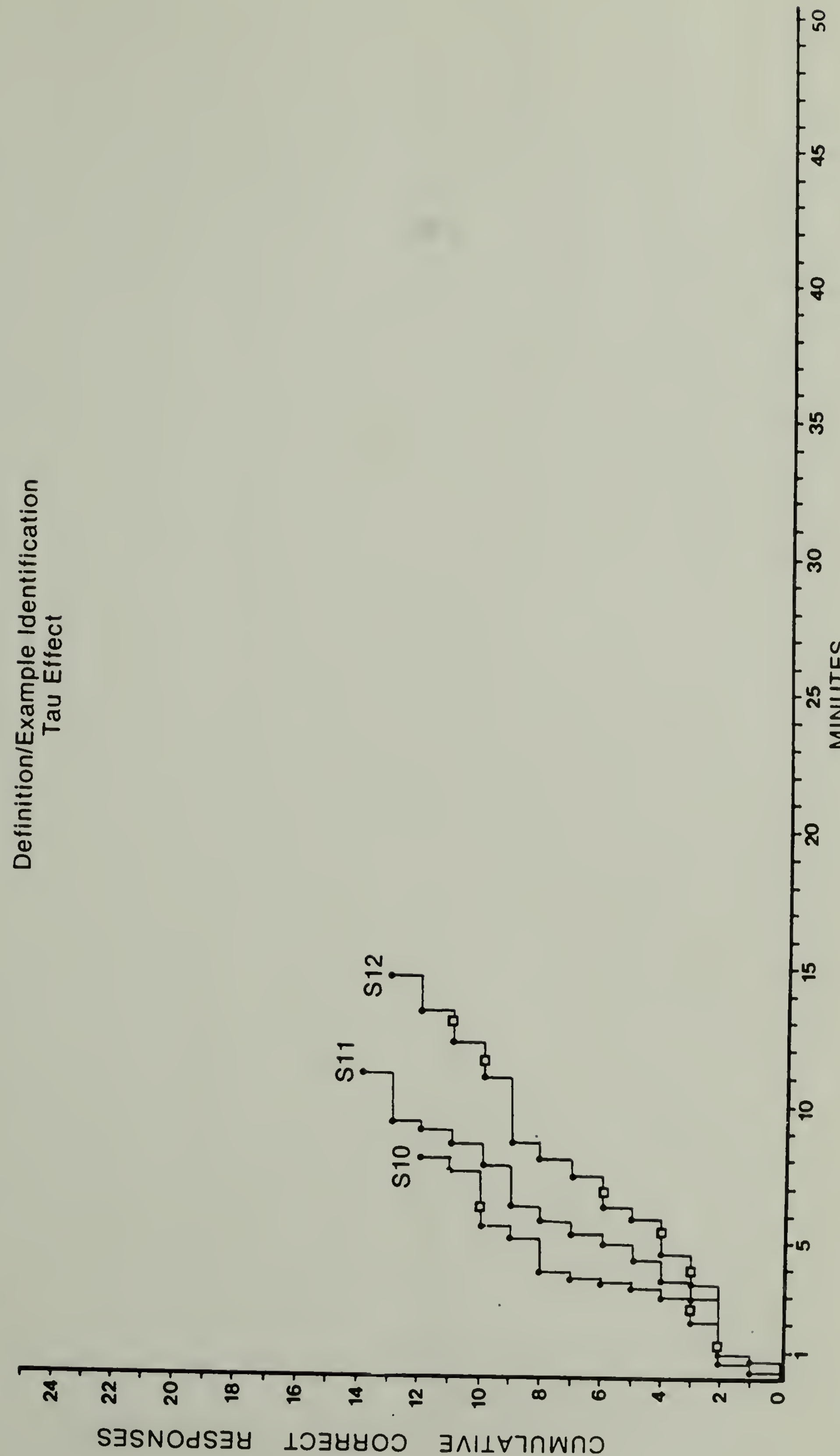
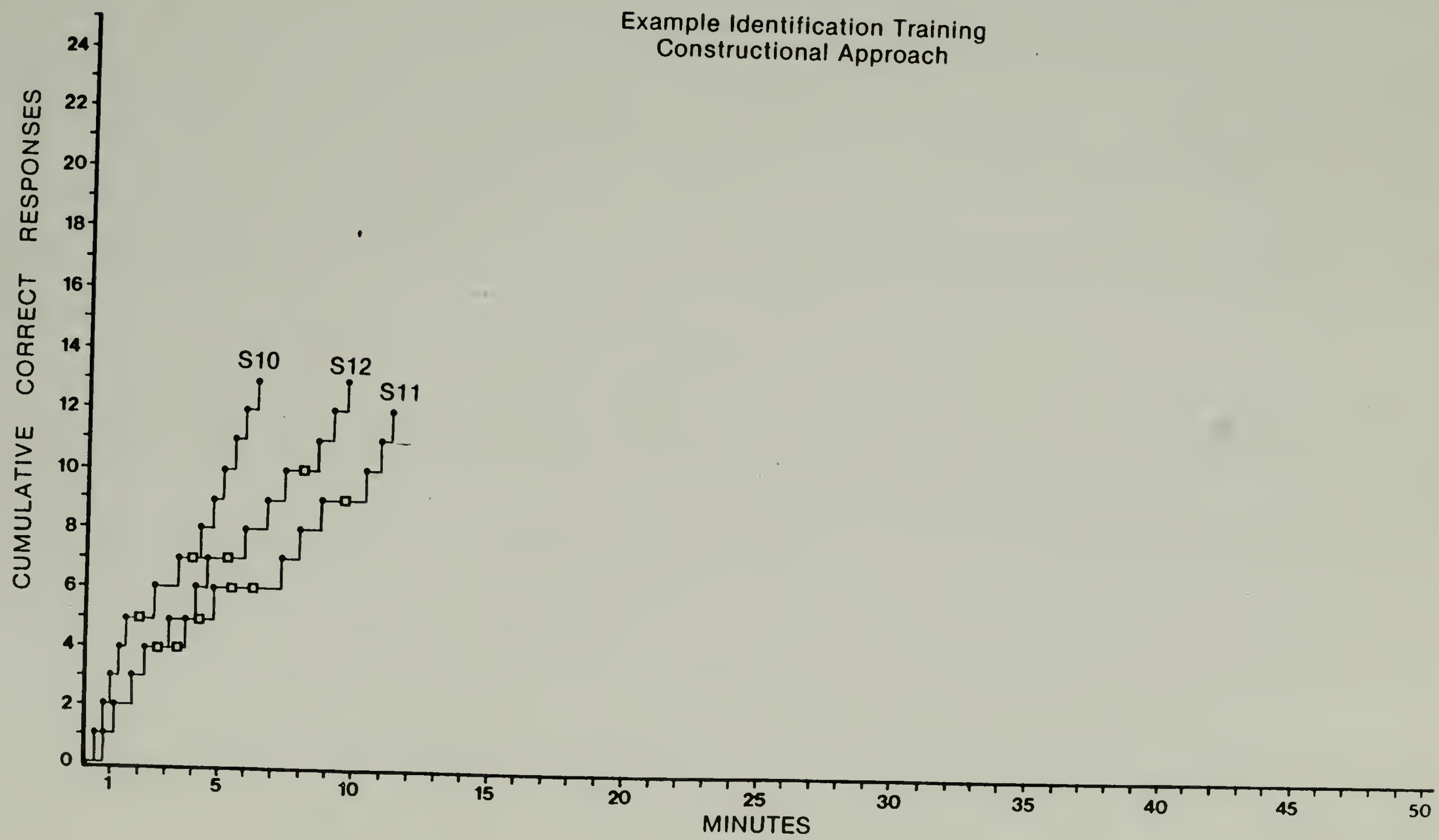


Figure 13. Cumulative frequency of correct study performance.


Definition/Example Identification
Tau Effect
CUMULATIVE CORRECT RESPONSES
0
2
4
6
8
10
12
14
16
18
20
22
24
MINUTES
1
5
10
15
20
25
30
35
40
45
50
S10
S11
S12


Example Identification Training
Constructional Approach
CUMULATIVE CORRECT RESPONSES
0
2
4
6
8
10
12
14
16
18
20
22
24
MINUTES
1
5
10
15
20
25
30
35
40
45
50
S10
S12
S11

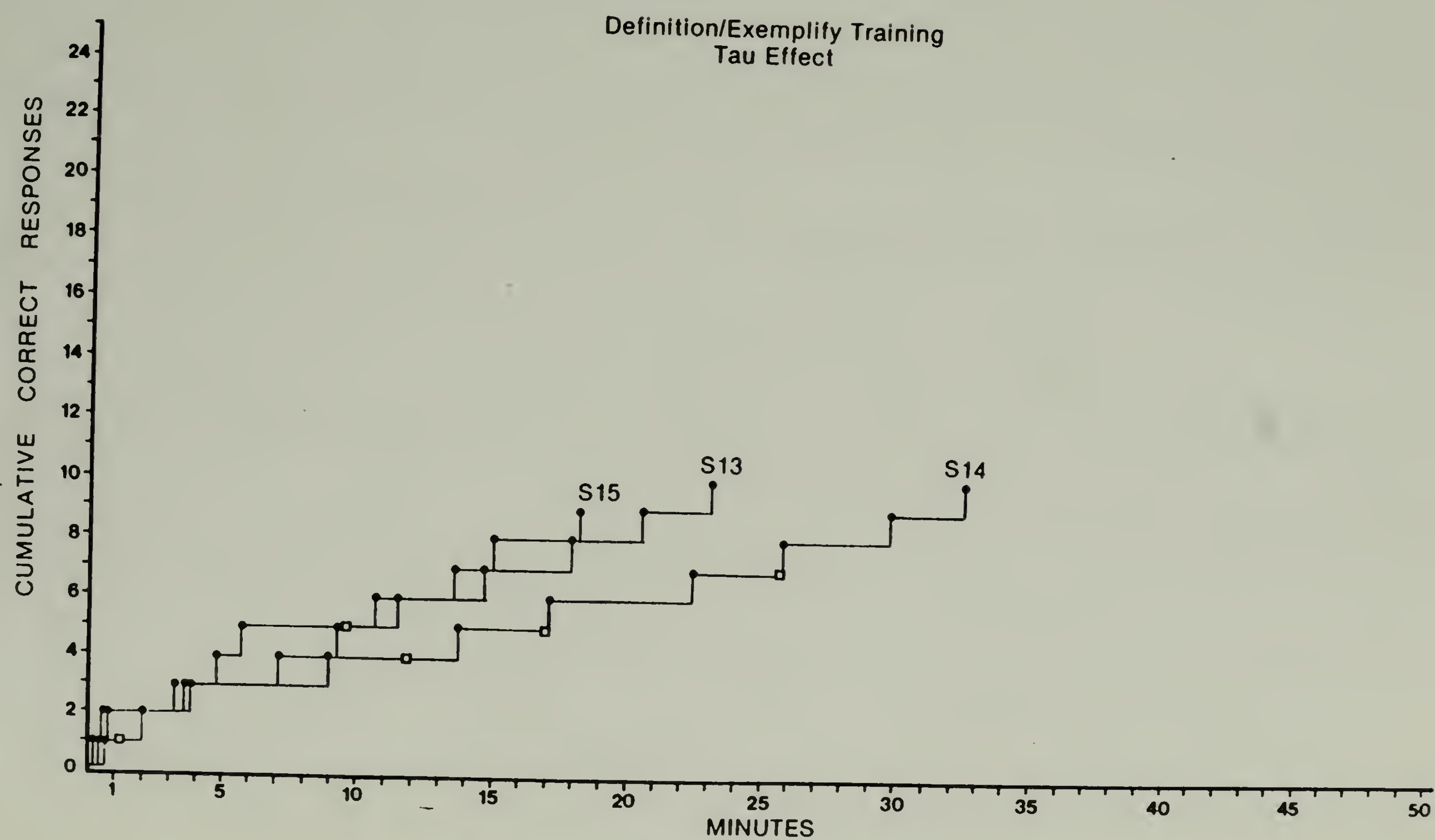
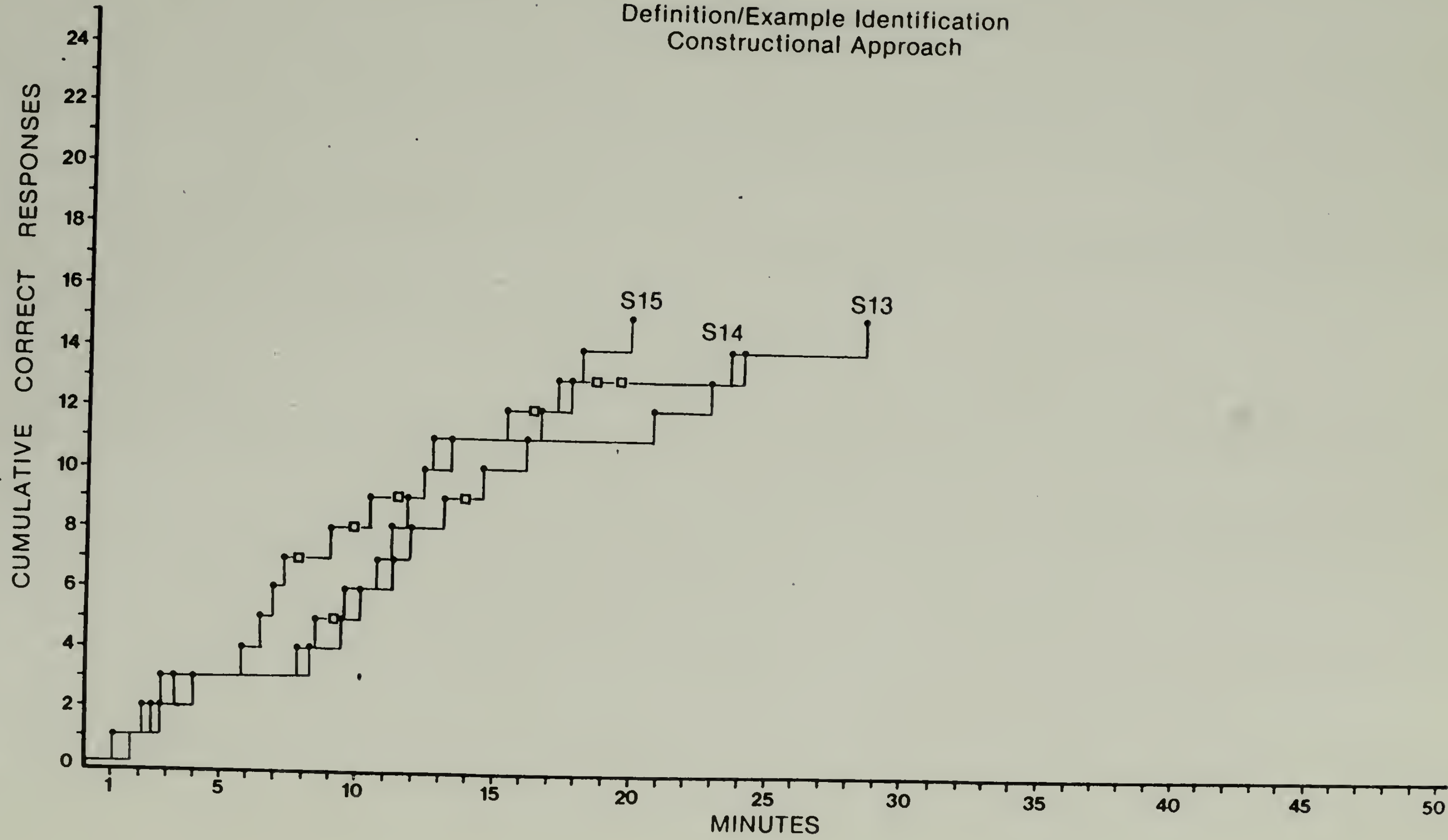
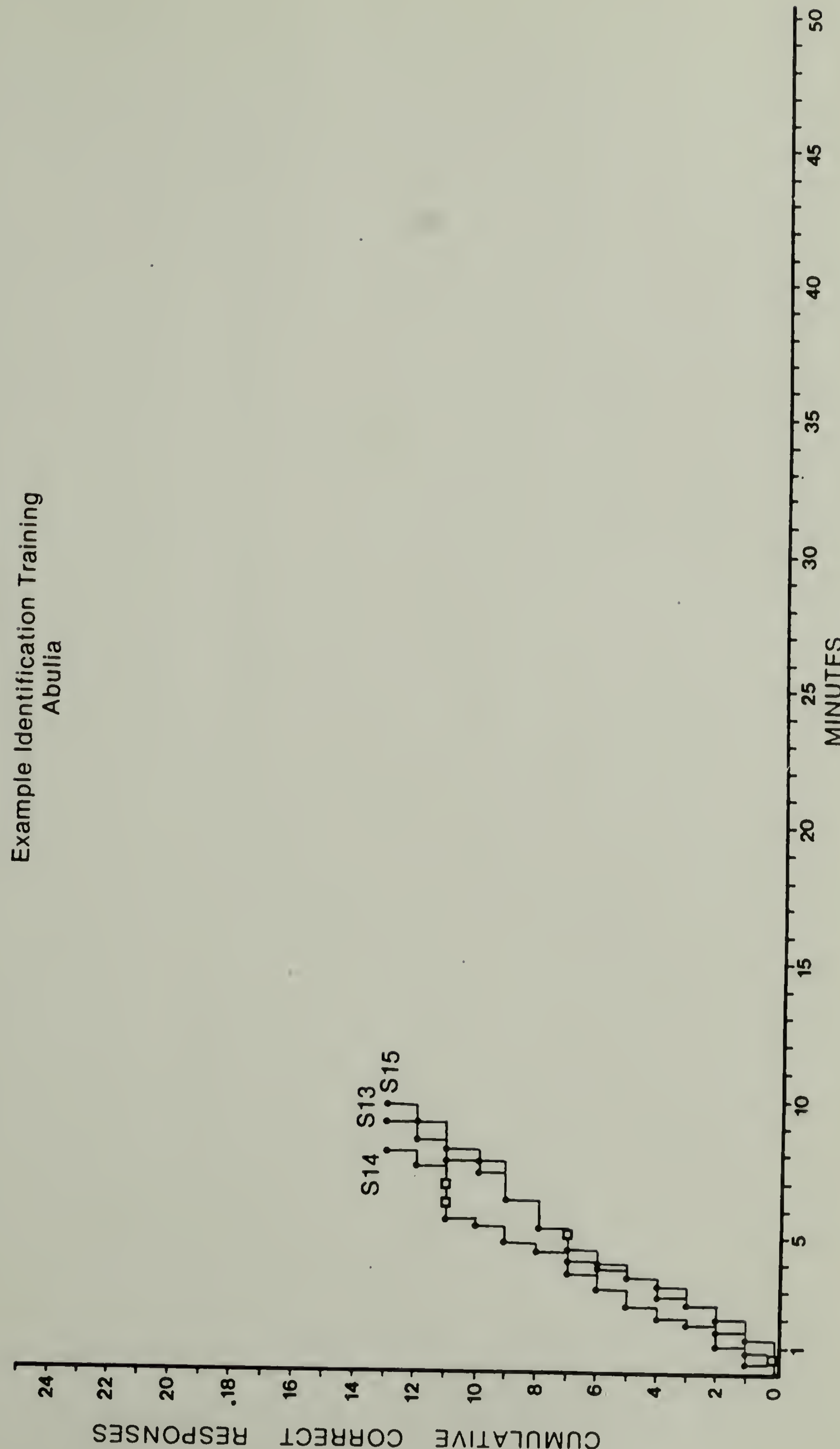


Figure 14. Cumulative frequency of correct study performance.

Definition/Example Identification
Constructional Approach




Example Identification Training
Abulia
CUMULATIVE CORRECT RESPONSES
0
2
4
6
8
10
12
14
16
.18
20
22
24
MINUTES
1
5
10
15
20
25
30
35
40
45
50
S14
S13
S15

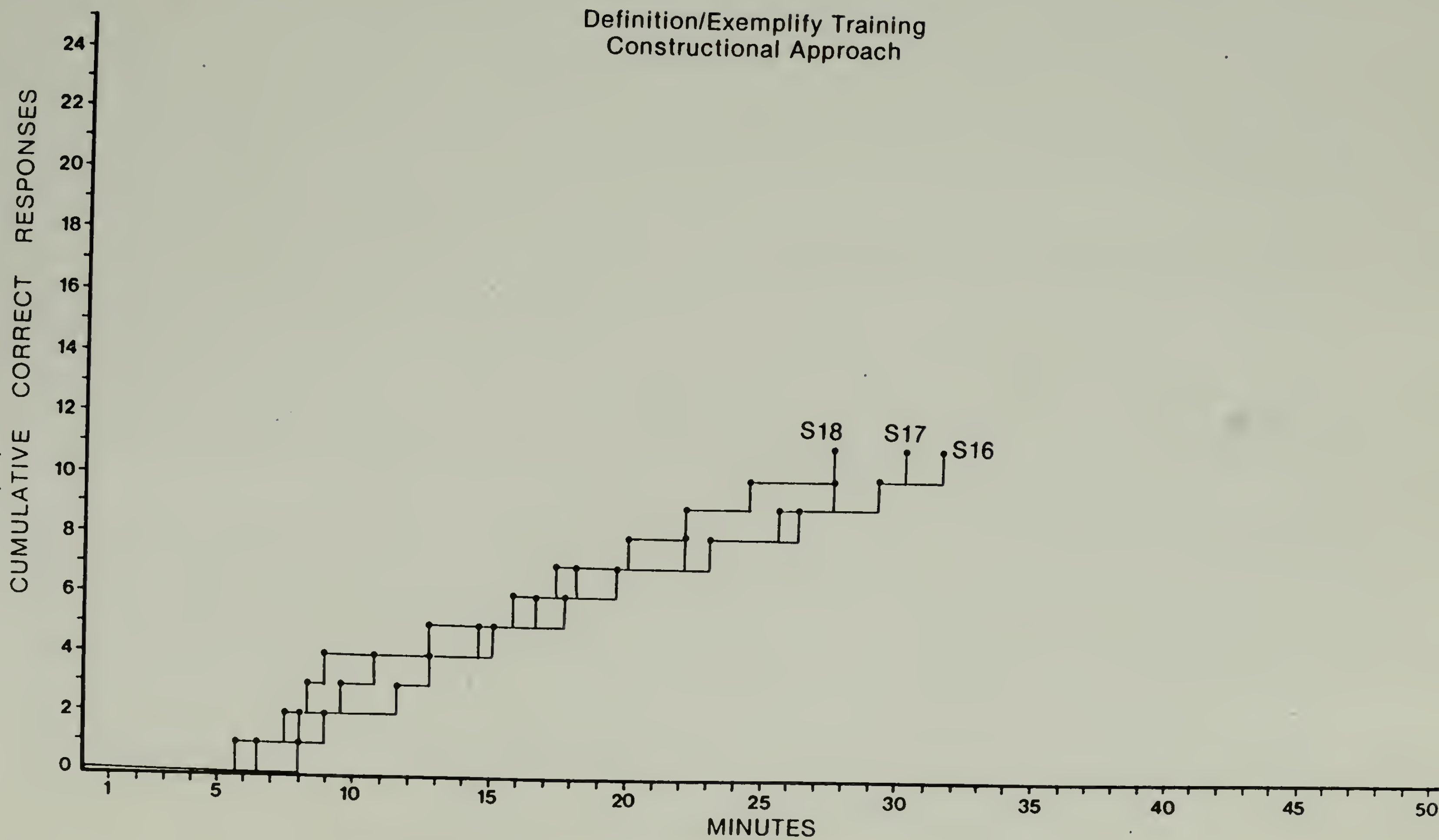
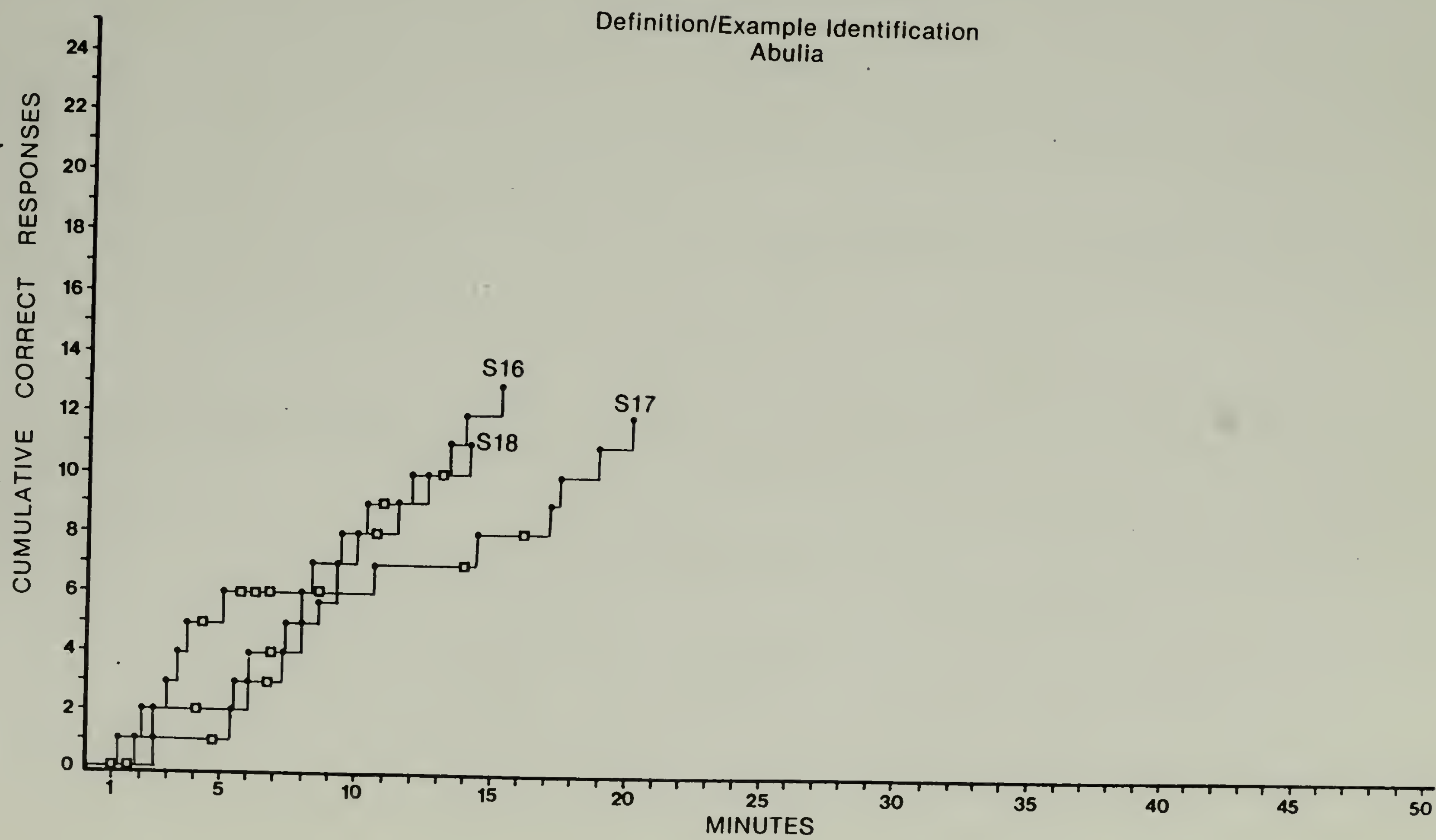
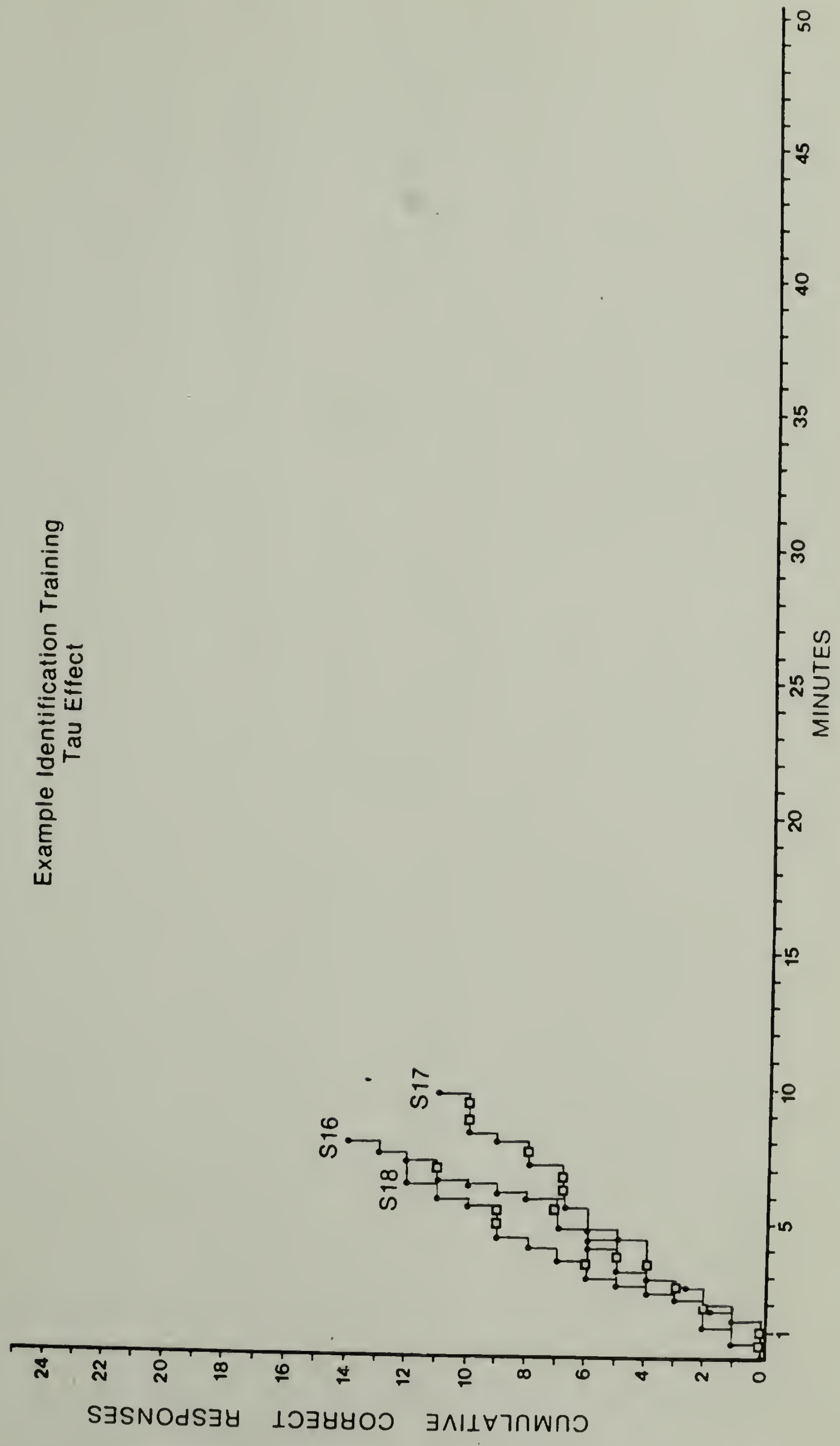


Figure 15. Cumulative frequency of correct study performance.


Definition/Example Identification
Abulia
CUMULATIVE CORRECT RESPONSES
0
2
4
6
8
10
12
14
16
18
20
22
24
1
5
10
15
20
25
30
35
40
45
50
MINUTES
S16
S17
S18


Example Identification Training
Tau Effect
CUMULATIVE CORRECT RESPONSES
MINUTES
S16
S17
S18
0
2
4
6
8
10
12
14
16
18
20
22
24
1
5
10
15
20
25
30
35
40
45
50

performance on example identification programs was higher than either definition/example identification programs or definition/exemplify programs. Rate of correct example identification performance was higher for 16 of the 17 subjects that could be compared. Subject 4's performance on the example identification program was terminated when he claimed that he was sick and could not continue with the session. The sole exception to this relation between example idenfification and other programs was Subject 8. Her rate of correct performance was slightly faster on the definition/example identification program than on the example identification program. A planned comparison between example identification programs and both definition/example identification and definition/exemplify programs yielded a significant difference, for rate $F\ (1,\ 24) = 72.04,\ p < .01$.

Second, rate of correct performance on definition/example identification programs was higher than definition/exemplify programs. Rate of correct responding on definition/example identification programs was higher for 13 of the 17 subjects that could be compared. Subject 2's performance on the definition/exemplify program was terminated when she did not answer the first four questions correctly. The four exceptions to this relation between definition/example identification and definition/exemplify programs (subjects 4, 5, 13 and 15) all received definition/example identification programs with the concept constructional approach. A planned comparison between definition/example identification and definition/exemplify programs also revealed a significant difference, $F\ (1, 24) = 42.22$, $p < .01$.

Figures 10-15 also revealed differences in duration of the study programs that can be attributed to both the study programs and the concepts. No differences were found to be related to either exposure to pretest or order of presenting the study programs.

First, example identification programs took less time to complete than either definition/example identification or definition/exemplify programs. Durations were shorter for 15 of the 17 subjects who could be compared. Both exceptions (Subjects 1 and 8) spent less time on definition/example identification programs, however, the differences were small. A planned comparison between example identification programs and both of the other programs reveal a significant difference, $F\ (1, 24) = 49.02$, $p < .01$.

Second, subjects took less time to complete definition/example identification programs than definition/exemplify programs. The durations were shorter for 13 of the 17 subjects who could be compared. All four exceptions (Subjects 4, 5, 13 and 15) occurred when subjects received definition/example identification programs with the concept *constructional approach*. A planned comparison between definition/example identification and definition/exemplify programs yielded a significant difference, $F\ (1,\ 24) = 28.08,\ p < .01$.

Third, duration measures also revealed a difference between concepts. Ten of the eighteen subjects spent more time to complete programs for the *constructional approach* than either *abulia* or *tau effect*. This difference was significant. A planned comparison between concepts revealed an $F\ (1,\ 24) = 19.48,\ p < .01$. However, no difference was found between *abulia* and *tau effect*.

Figures 16 and 17 present the percent correct performance on the study program for all eighteen subjects. Figure 17 presents the data for the nine subjects who received a pretest. Figure 18 presents the data for the nine subjects who did not receive a pretest. On both graphs subjects are grouped according to the combination of study program and concept that they received. Open bars represent performance on example identification programs, solid bars illustrate performance on definition/exemplify programs. Symbols at the bottom of each graph indicate the concept received (A = *abulia*, C = *constructional approach* and T = *tau effect*). These data reveal systematic differences related to study program. However, no differences were found related to pretest, order of study programs or concepts.

- - - - - - - - - - - - - - - - - - - -

SEE FIGURES 16 & 17, PAGES 165-166

- - - - - - - - - - - - - - - - - - - -

Specifically, percent correct performance on definition/exemplify programs was higher than on both example identification programs and definition/example identification programs. Definition/exemplify performance was higher than example identification performance for 13 of the 16 subjects who could be compared. For one exception (Subject 5) the difference was 2% points, the other two exceptions had equal performance on define exemplify and example identification (Subjects 13 and 15). Definition/exemplify performance was higher than definition/example identification performance for 11 of the 17 subjects who could be compared. Two of the six exceptions (Subjects 8 and 17) had equal performance on the two study programs. A planned comparison between definition/exemplify programs and the other programs reveal a significant difference, F 91, 24) = 39.74, $p < .01$. The difference between definition/example identification and example identification programs was not significant.

Further group analyses were conducted to substantiate the comparisons described above. A four way repeated measures, latin square analysis of variance (ANOVA) was conducted for each dependent measure. Specifically, three 2 (pretest) x 3 (order of program presentation) x 3 (concept)

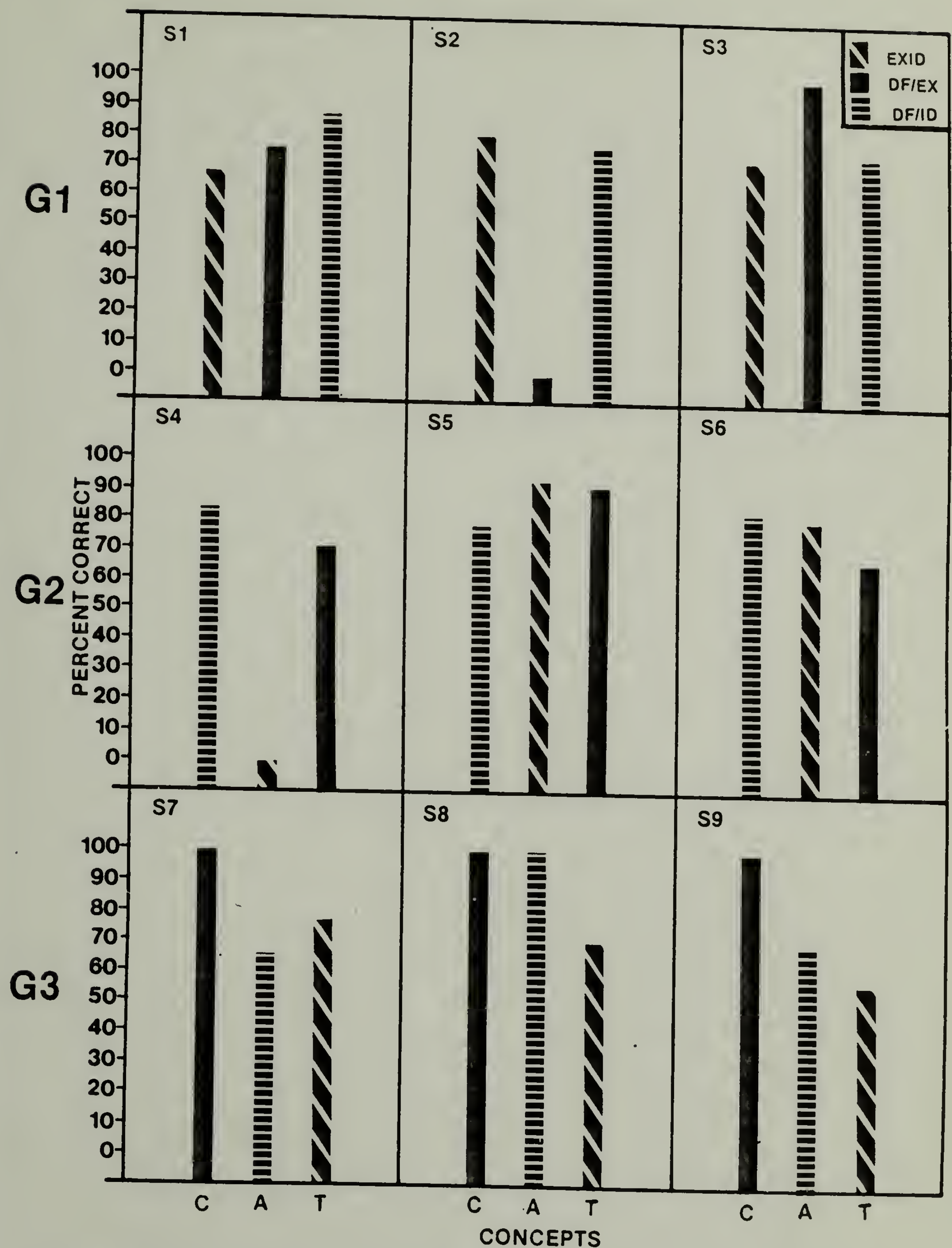


Figure 16. Percent correct on study programs for the subjects who received a pretest.

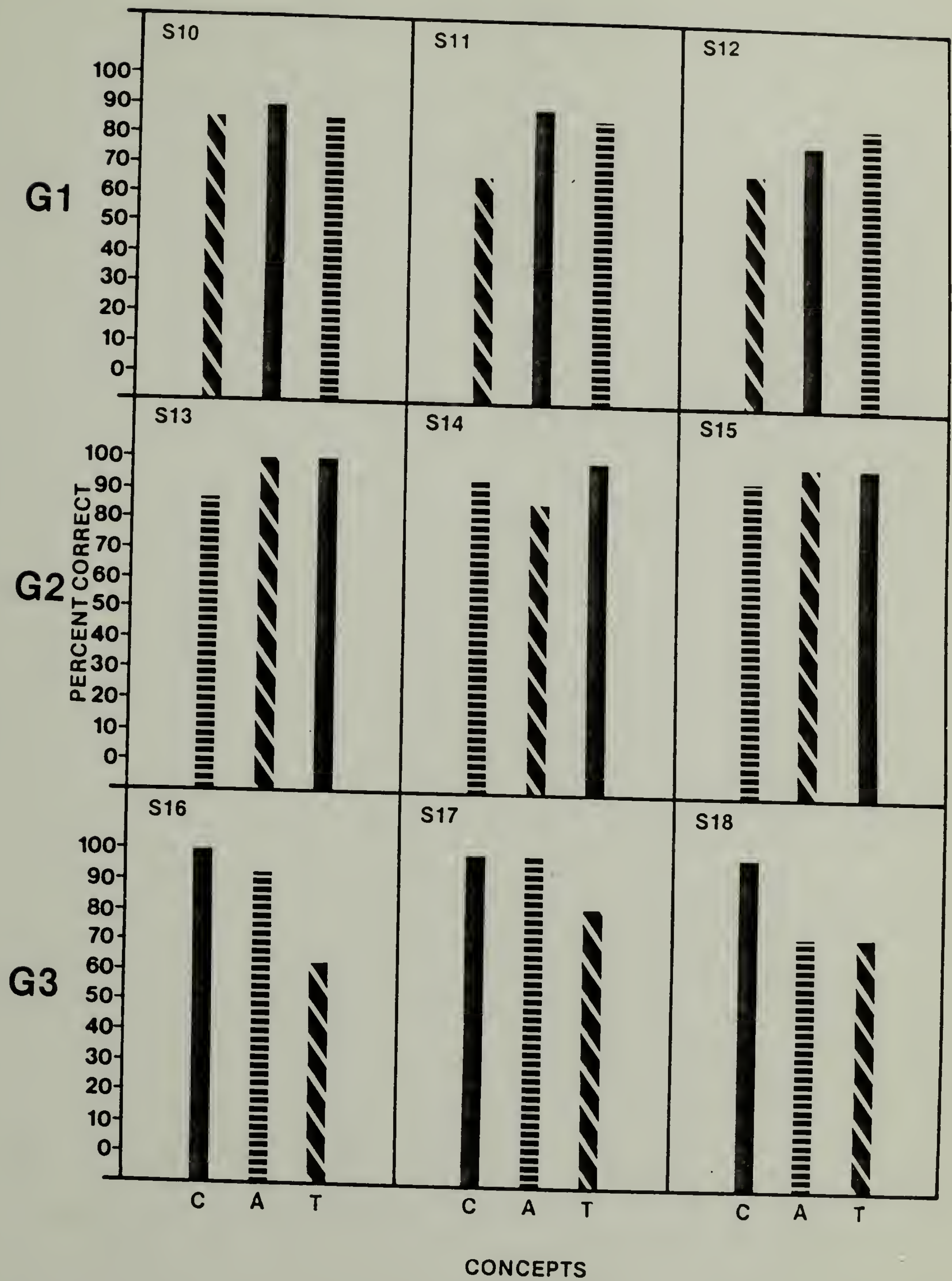


Figure 17. Percent correct on study programs for subjects who did *not* receive a pretest.

x 3 (study program (ANOVA's were calculated. An arc sin transformation of proportions was calculated for all percent correct data to stabilize the variance in order to assume homogeneity of variance.

Significant effects of pretest, or order of presenting the study programs were not found on any of the dependent measures of study performance. In addition, estimates of the interaction between order and concept and the interaction between pretest, order and concept were not significant. The absence of these interactions indicates that the tests for the main effects of concept and study program were unbiased.

ANOVA's for all three dependent measures yielded significant effects of type of study program. Table 17 presents the source data for the ANOVA for correct rate of responding. The only significant effect found was for study program. The study program effects were obtained by partitioning the order of study program by concept interaction into study program and residual effects. Since there was no order effects nor a residual effect, the F test for study program was unbiased. The F test for study program was significant, $F\ (2, 24) = 57.23$, $p < .01$. Tables 18 and 19 present similar analyses for both percent correct performance and duration of study session respectively. The F ratios for study program were significant in both, $F\ (2, 24) = 20.84$, $p < .01$ and $F\ (2, 24) = 338.74$, $p < .01$.

respectively. In addition, Table 18 indicates an interaction between pretest and order on percent correct performance, $F (2, 12) = 4.63$, $p. < .05$. Table 19 indicates that there is a significant effect of concept on duration measures, $F (2, 24) = 9.24$, $p < .01$.

- - - - - - - - - - - - - - - - - - -

SEE TABLES 17, 18, 19, PAGES 169-171

- - - - - - - - - - - - - - - - - - -

In summary, group and intrasubject analyses of study performance both revealed systematic differences as a function of study program on all dependent measures. It appears that example identification study programs take less time to complete and the questions are responded to faster, but performance is less accurate than either of the other two programs. Definition/exemplify programs appear to facilitate accuracy. However, they take significantly more time to complete than either definition/example identification or example identification programs. In addition, there appears to be a difference between concepts on duration measures. Constructional approach programs take longer to complete than either abulia or tau effect programs. Finally, group analyses of percent correct performance revealed an interaction between pretest and order of study program. Apparently, subjects who received the third order of training (Define/Example ID, Example ID, Define/Exemplify) were more affected by whether they received a pretest than

TABLE 17

ANALYSIS OF VARIANCE FOR RATE OF CORRECT RESPONDING DURING STUDY SESSIONS

Source	Sum of Squares	DF	Mean Squares	F
Mean	26.418	1	26.418	356.31**
Pretest (A)	.012	1	.012	.16
Order (B)	.278	2	.138	1.71
AB	.123	2	.061	.75
Error	.971	12	.081	
Concept (C)	.041	2	.020	.481
AC	.092	2	.045	.205
BC				
Program (D)	3.099	2	1.548	57.23**
Residual	.085	2	.043	1.57
ABC				
AD	.126	2	.063	2.33
Residual	.151	2	.075	2.79
Error	.649	24	.027	

** indicates $p < .01$

TABLE 18

ANALYSIS OF VARIANCE FOR PERCENT CORRECT PERFORMANCE ON STUDY PROGRAMS

Source	Sum of Squares	DF	Mean Squares	F
Mean	338580.33	1	338580.33	2765.30**
Pretest (A)	113.337	1	113.337	.93
Order (B)	608.691	2	304.598	2.49
AB	1133.197	2	566.598	4.63*
Error	1224.388	12	102.032	
Concept (C)	204.452	2	102.226	1.91
AC	102.452	2	51.146	.95
BC				
Program (D)	1863.44	2	931.72	20.84**
Residual	200.18	2	100.10	2.24
ABC				
AD	303.64	2	151.82	3.39
Residual	159.90	2	79.95	1.78
Error	1072.44	24	44.69	

* indicates $p < .05$

** indicates $p < .01$

TABLE 19

ANALYSIS OF VARIANCE FOR DURATION OF STUDY PROGRAM

Source	Sum of Squares	DF	Mean Squares	F
Mean	23848.032	1	23848.032	509.68**
Pretest (A)	1.218	1	1.218	.03
Order (B)	329.675	2	164.837	3.42
AB	23.296	2	11.648	.25
Error	561.487	12	46.790	
Concept (C)	443.136	2	221.568	9.24**
AC	10.791	2	5.395	.24
BC				
Program(D)	1728.130	2	864.06	338.74**
Residual	26.86	2	13.43	.60
ABC				
AD	13.75	2	6.87	.30
Residual	34.90	2	17.45	.78
Error	535.240	24	22.30	

* indicates $p < .05$

** indicates $p < .01$

those subjects who received either the first or second order.

Test performance. Total test performance was analyzed with three kinds of data: rate of correct responding, percent correct and duration. Intrasubject analyses of these data revealed several relations. Planned comparisons of the variables were conducted to substantiate these analyses. In addition, two degrees of freedom were lost for each analysis because the data for two subjects were estimated. A Mahalanobis D-Squared and squared Multiple Correlation with all available data were used to estimate the missing data (Health Sciences Computing Facility, 1977).

Figures 18 and 23 present the data for correct responses per minute (rate) on the total test for all eighteen subjects. Graphing conventions are similar to those presented earlier (Figures 10-15). Subjects are grouped according to the combination of study program and concept that they received. The symbols C, A, and T represent constructional approach, abulia and tau effect respectively. Data presented on Figures 18-20 are for those subjects who received a pretest and the data presented on Figures 21-23 are for subjects who did not receive a pretest.

- - - - - - - - - - - - - - - - - -

SEE FIGURES 18-23, PAGES 173-190

- - - - - - - - - - - - - - - - - -

First, systematic differences between study programs

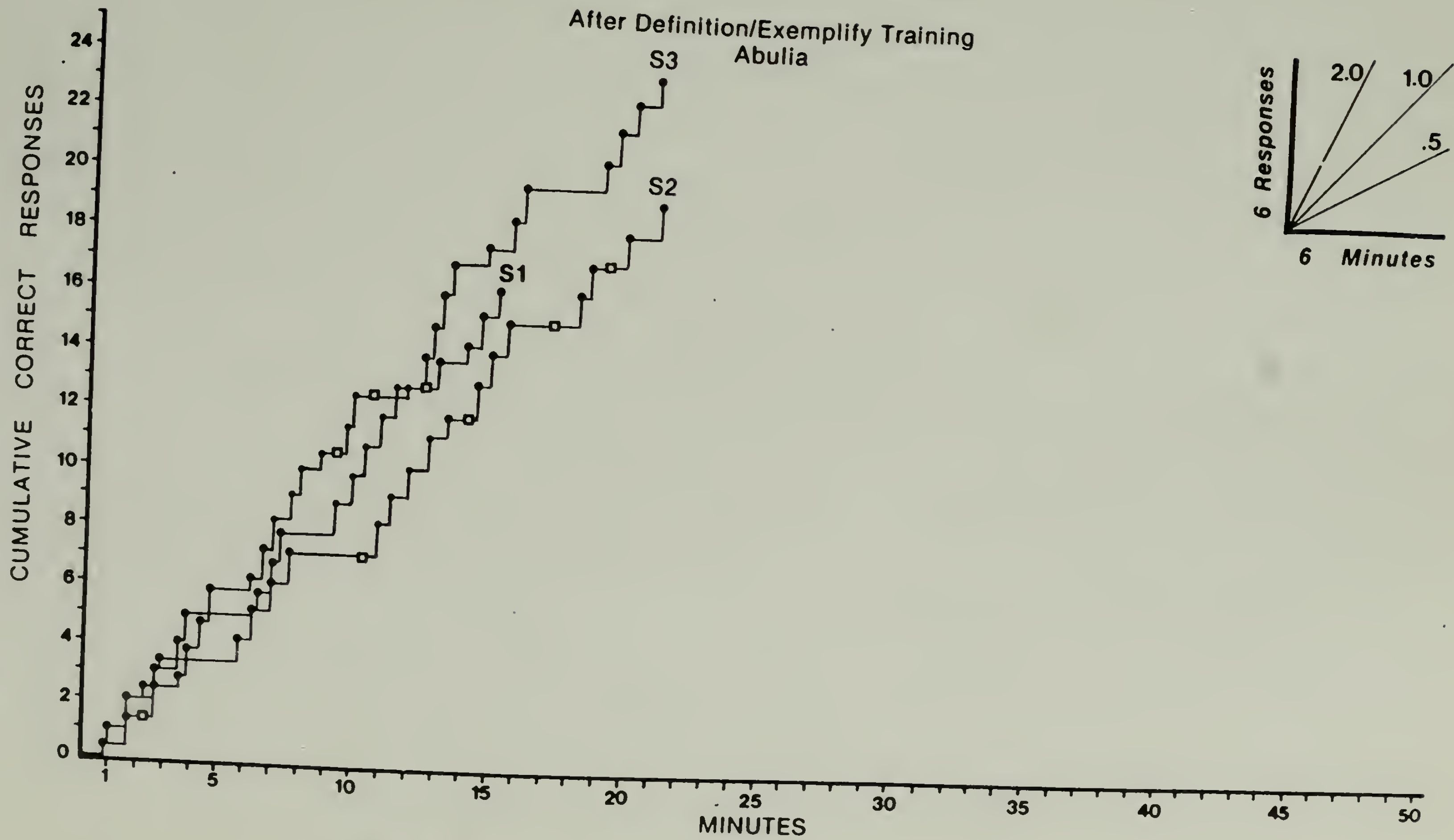
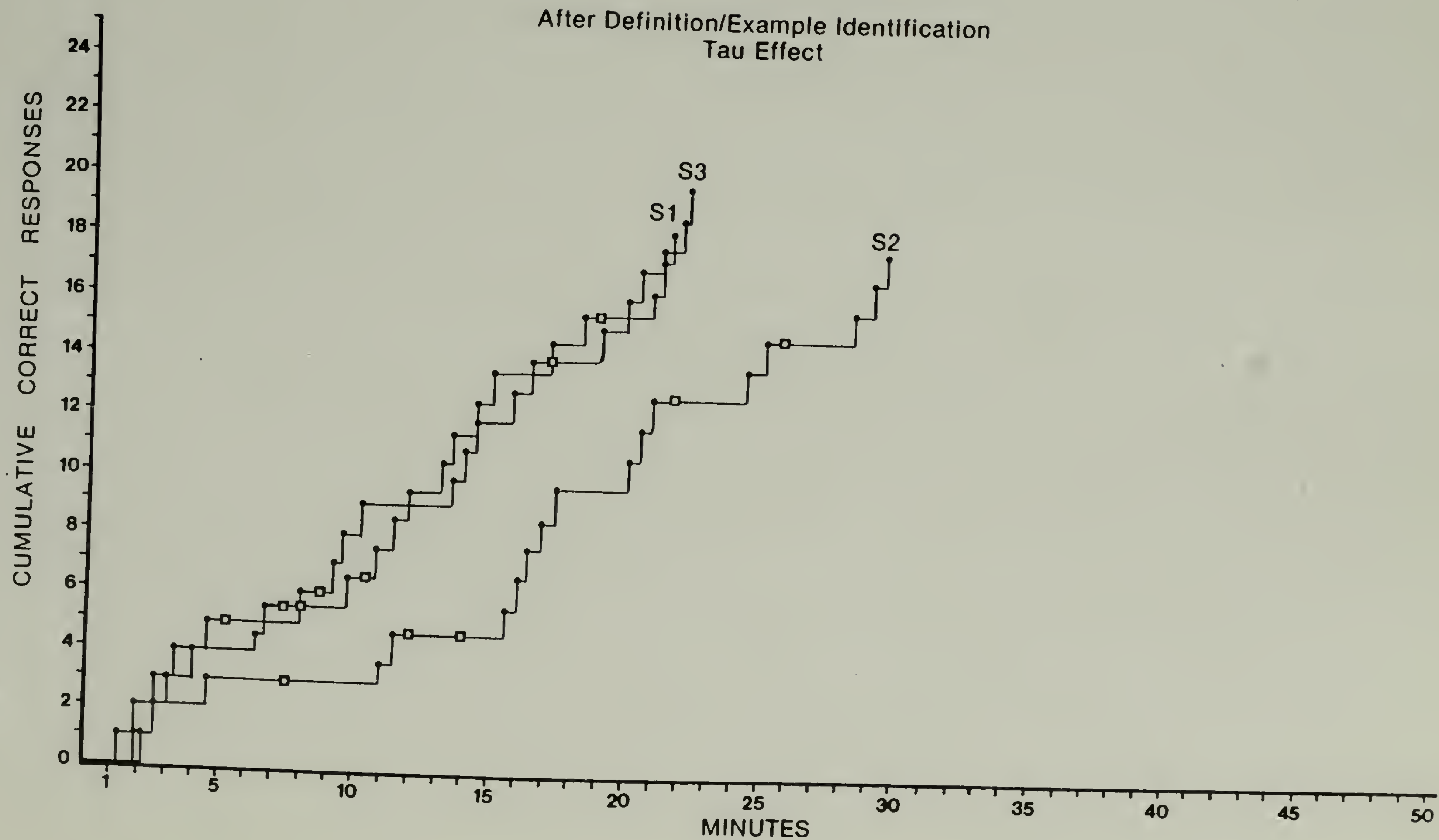
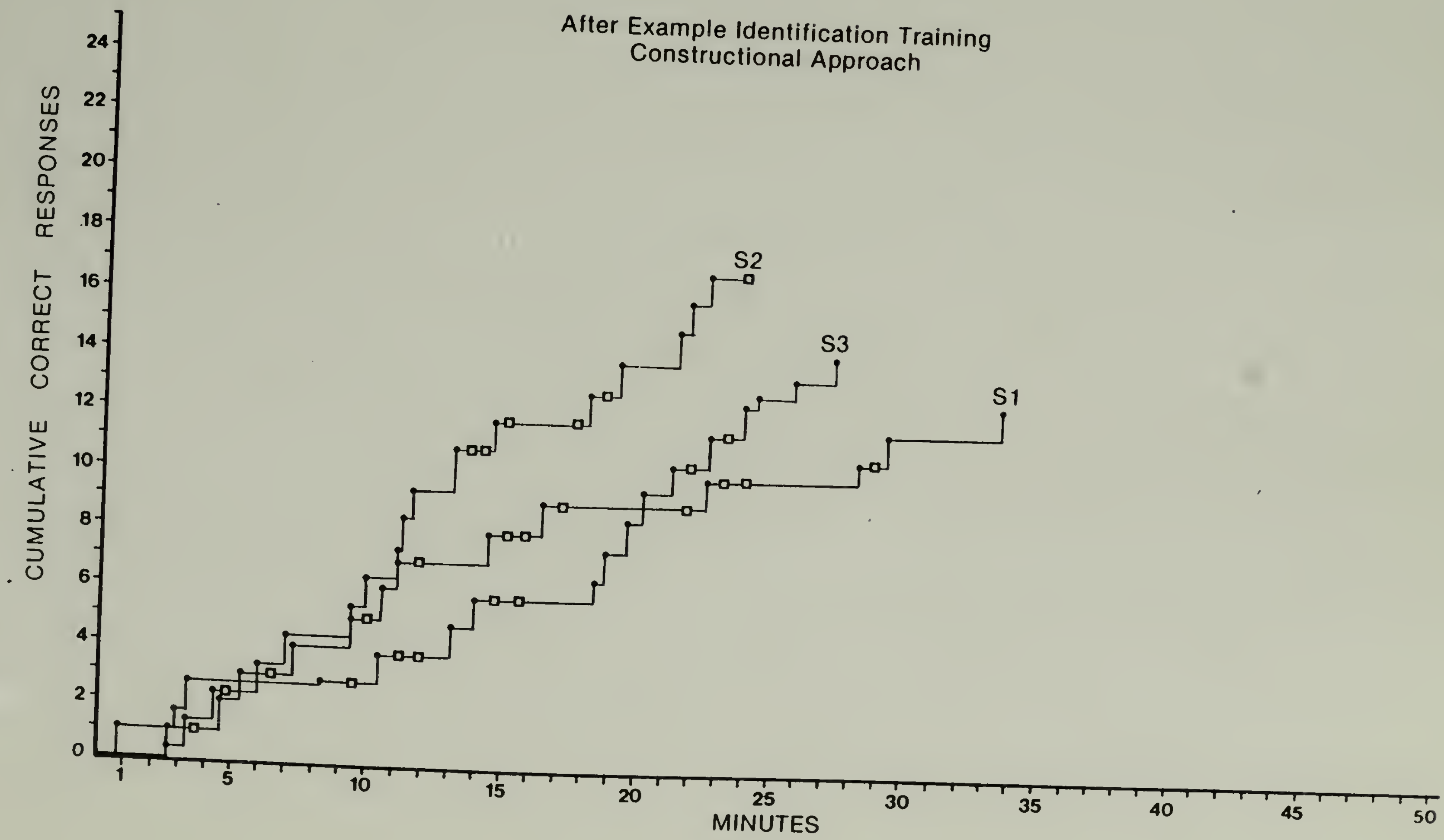


Figure 18. Cumulative frequency of correct test performance.


After Definition/Example Identification
Tau Effect
CUMULATIVE CORRECT RESPONSES
0
2
4
6
8
10
12
14
16
18
20
22
24
S1
S2
S3
1
5
10
15
20
25
30
35
40
45
50
MINUTES


After Example Identification Training
Constructional Approach
CUMULATIVE CORRECT RESPONSES
MINUTES
S1
S2
S3

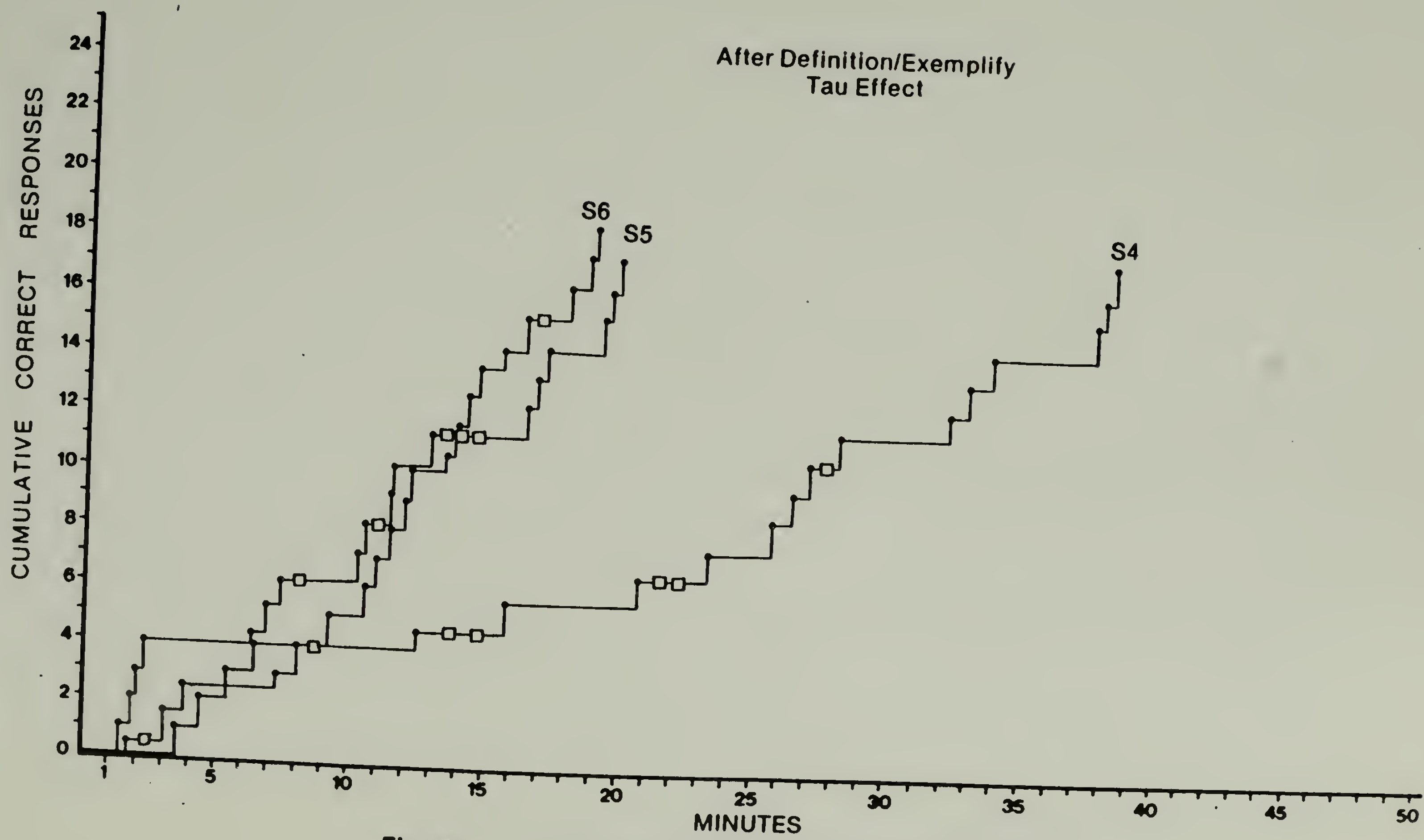
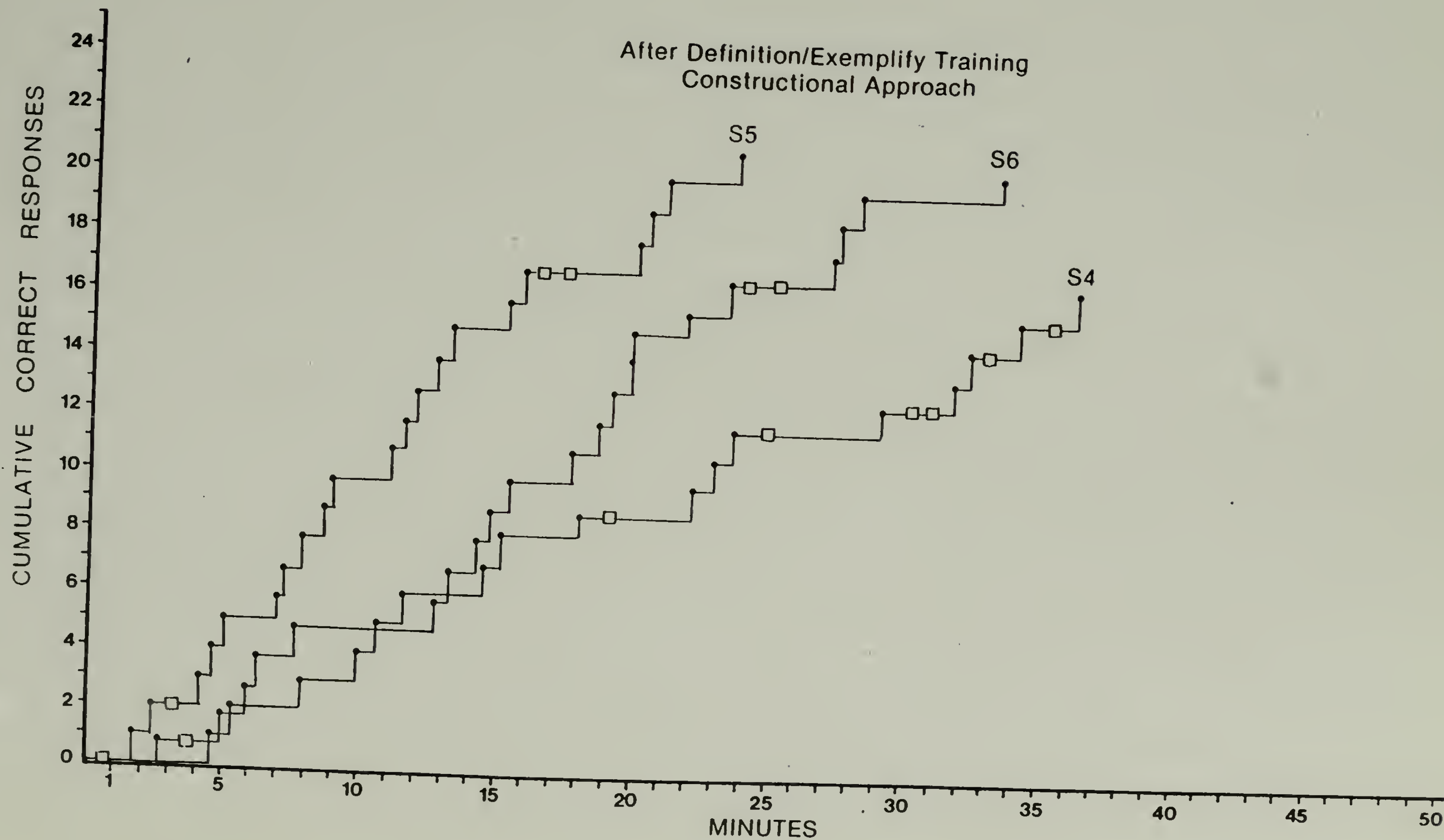
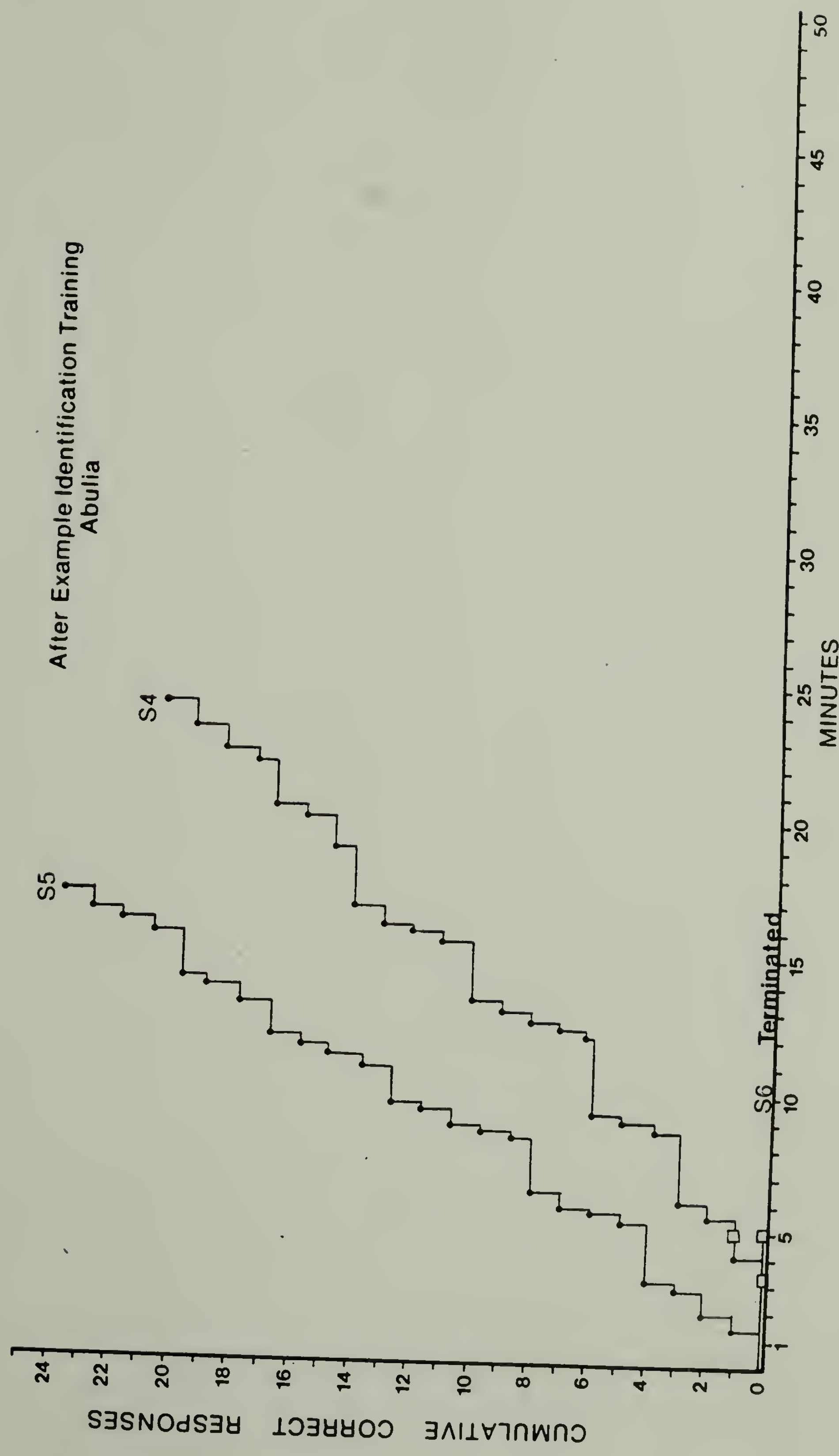


Figure 19. Cumulative frequency of correct test performance.


After Definition/Exemplify Training
Constructional Approach
CUMULATIVE CORRECT RESPONSES
MINUTES
S5
S6
S4
0
2
4
6
8
10
12
14
16
18
20
22
24
1
5
10
15
20
25
30
35
40
45
50



After Example Identification Training
Abulia
S5
S4
S6
Terminated
CUMULATIVE CORRECT RESPONSES
MINUTES
0
2
4
6
8
10
12
14
16
18
20
22
24
1
5
10
15
20
25
30
35
40
45
50

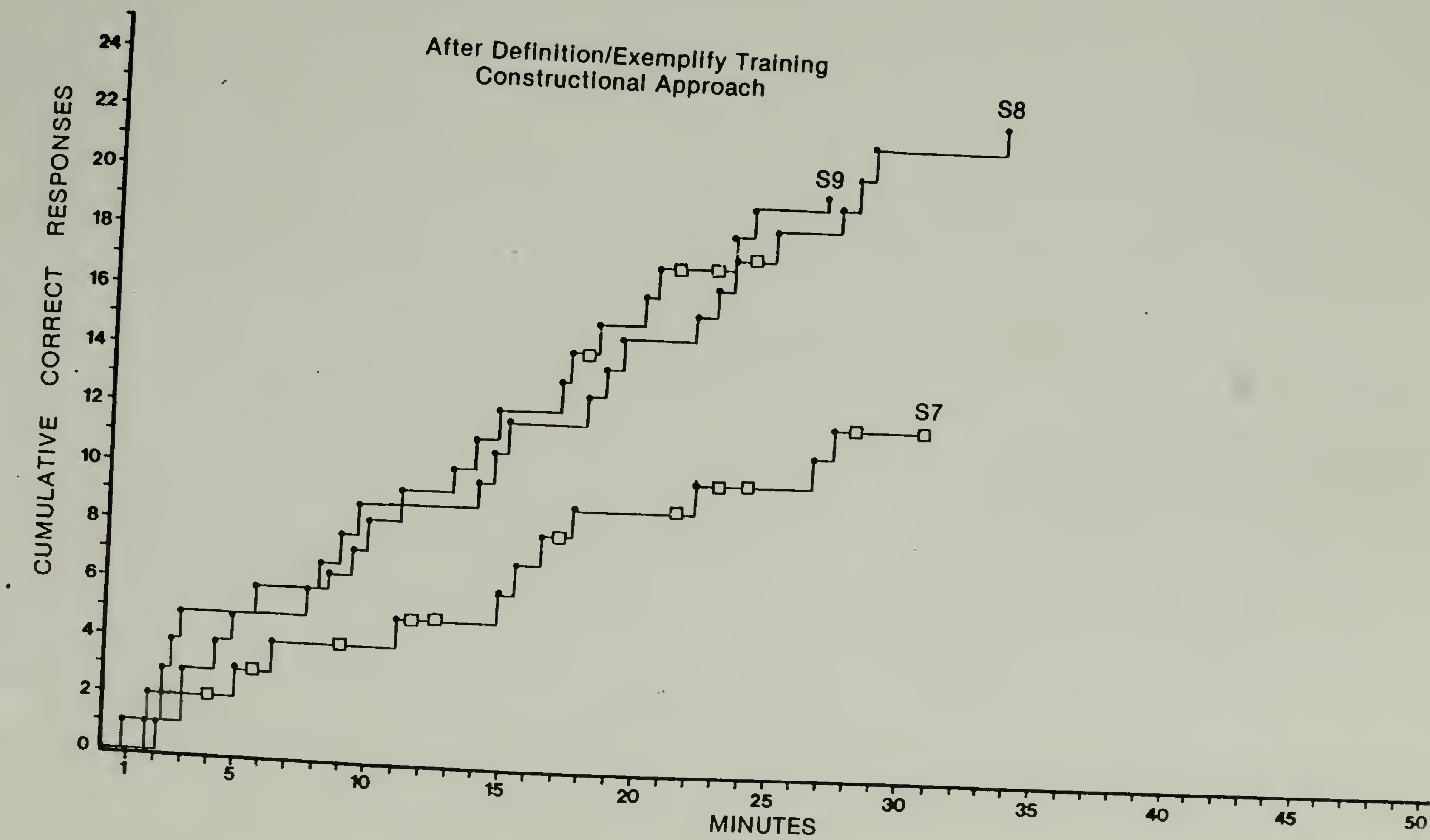
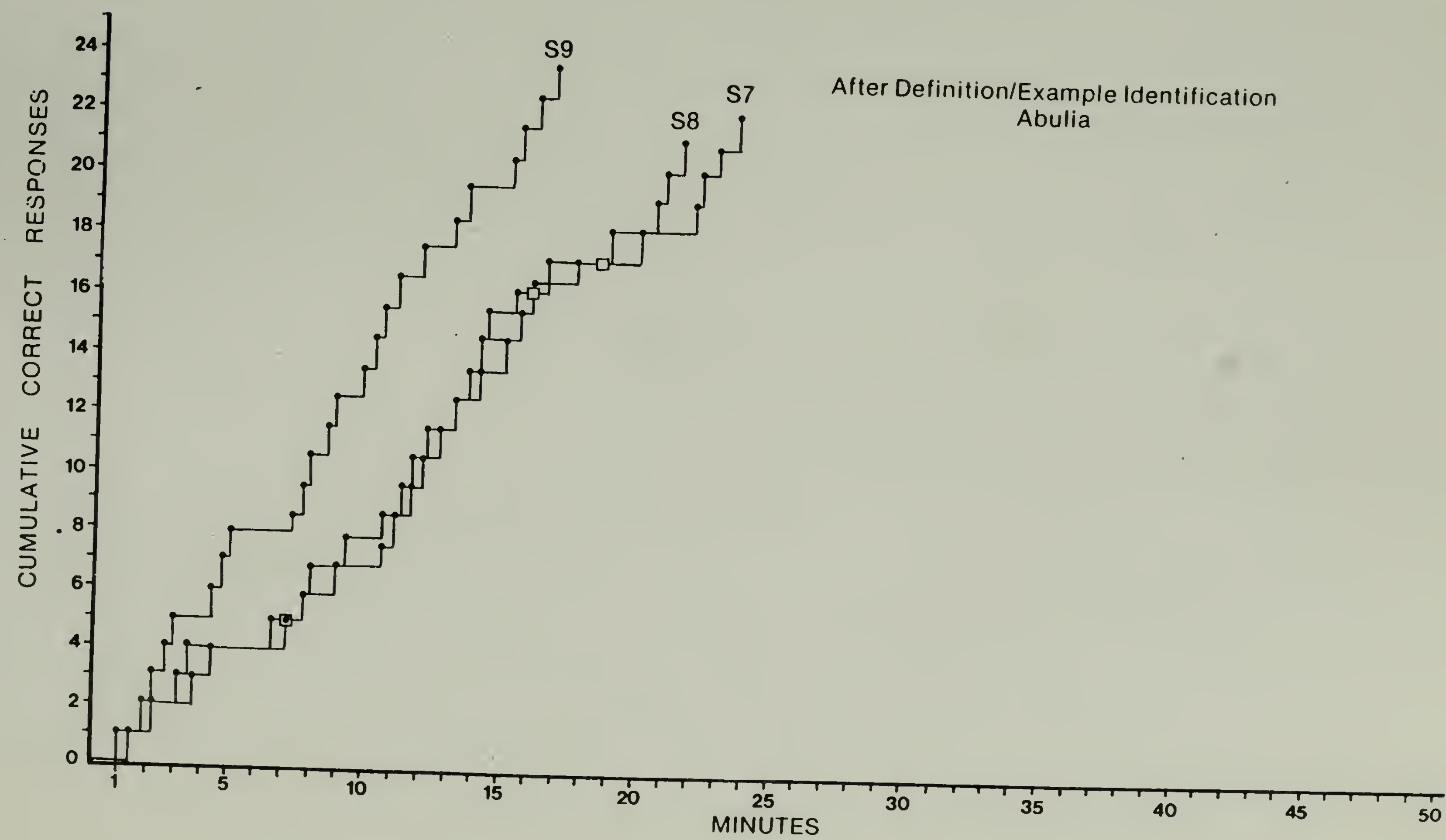
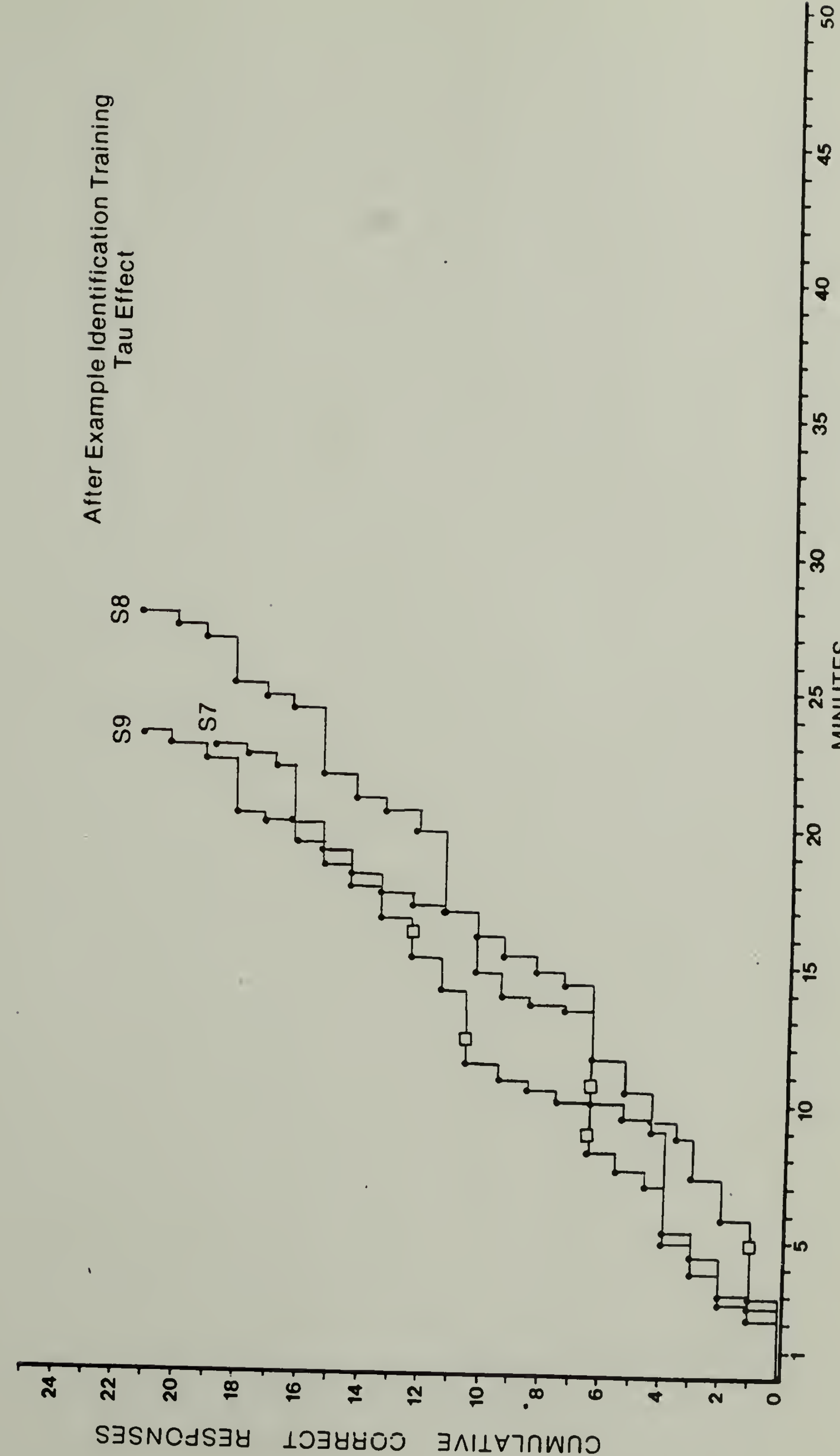


Figure 20. Cumulative frequency of correct test performance.


After Definition/Example Identification
Abulia
S9
S8
S7
CUMULATIVE CORRECT RESPONSES
0
2
4
6
8
10
12
14
16
18
20
22
24
1
5
10
15
20
25
30
35
40
45
50
MINUTES


After Example Identification Training
Tau Effect
S9
S7
S8
CUMULATIVE CORRECT RESPONSES
MINUTES
0
2
4
6
8
10
12
14
16
18
20
22
24
1
5
10
15
20
25
30
35
40
45
50

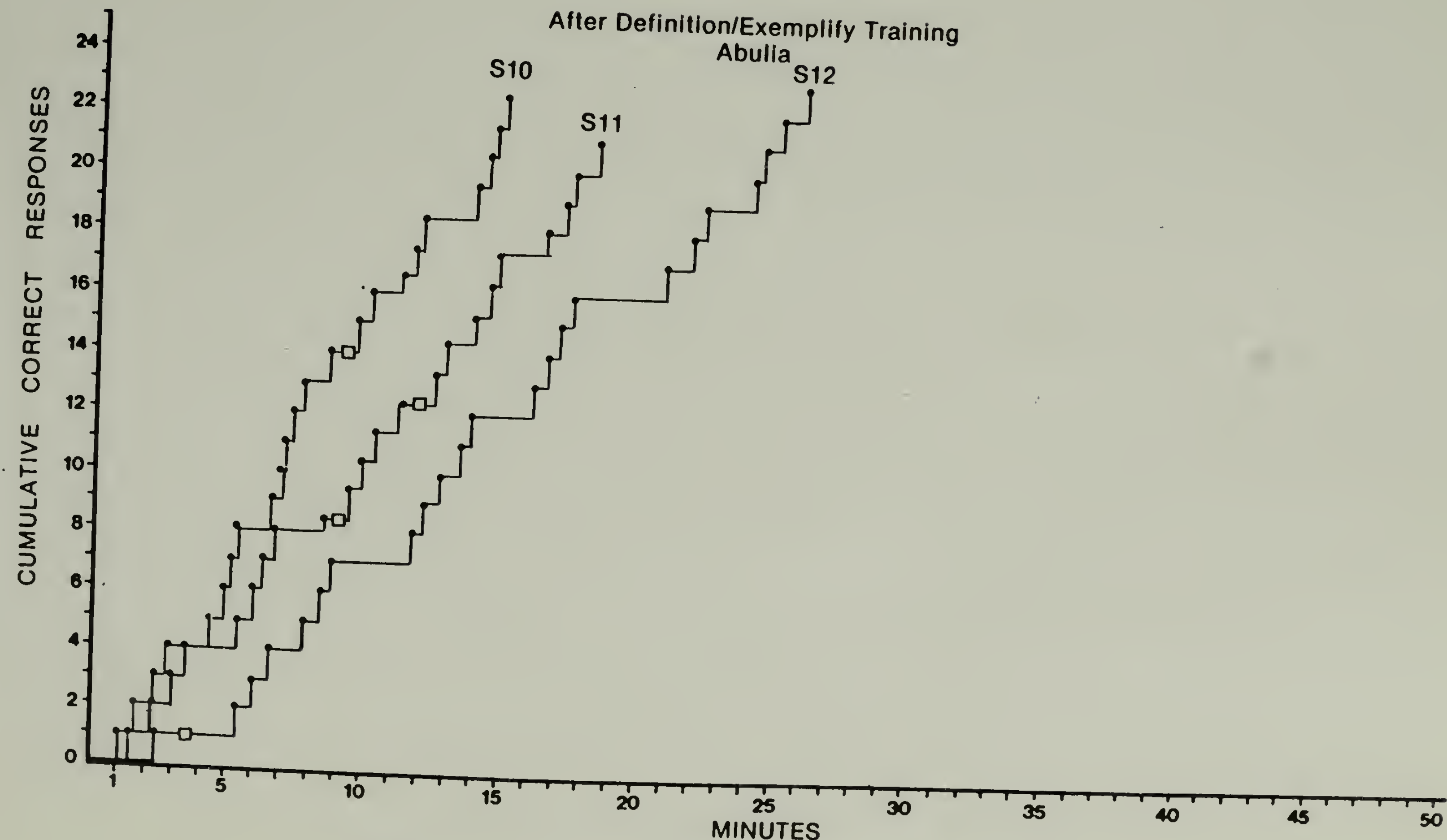
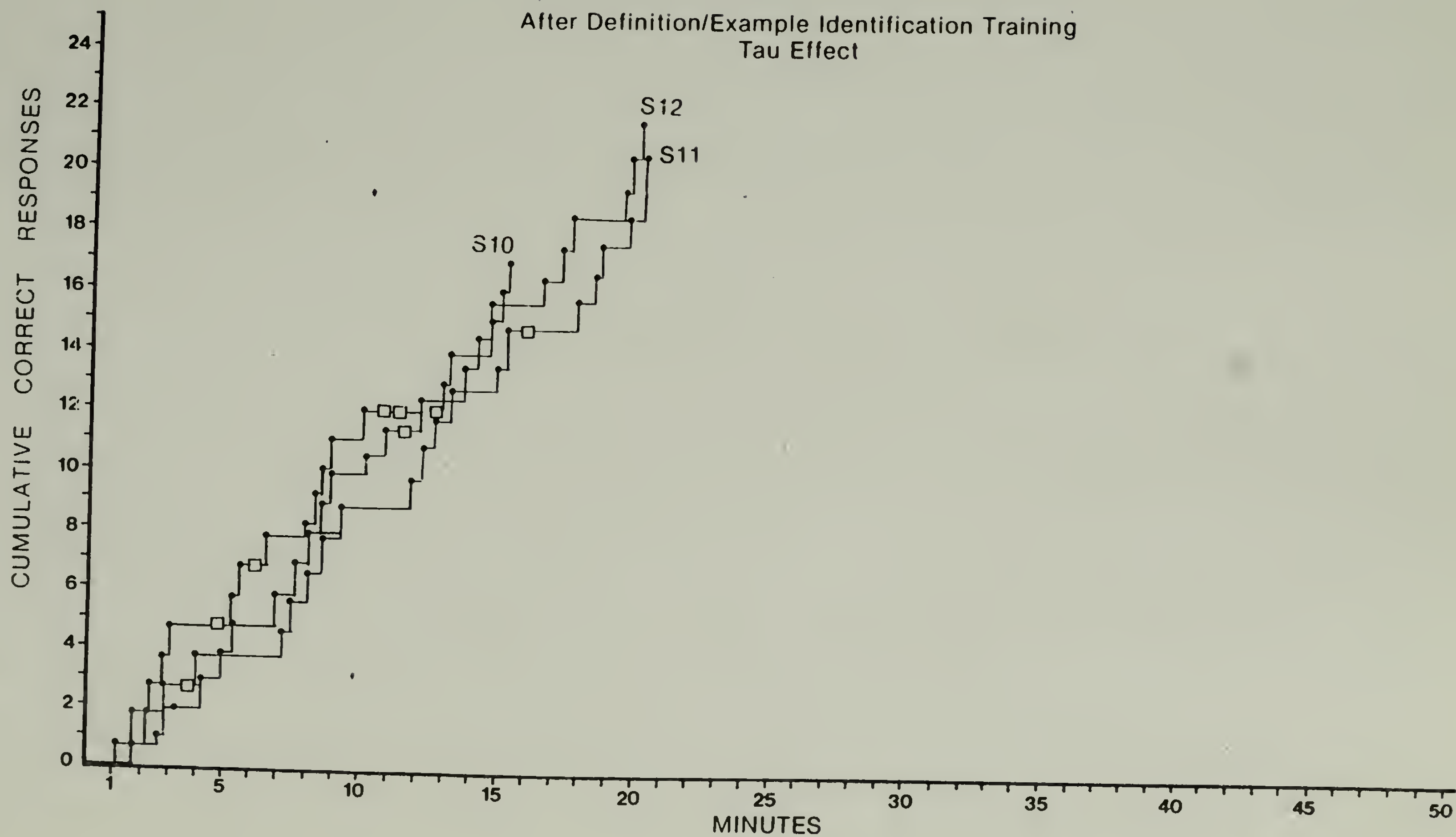
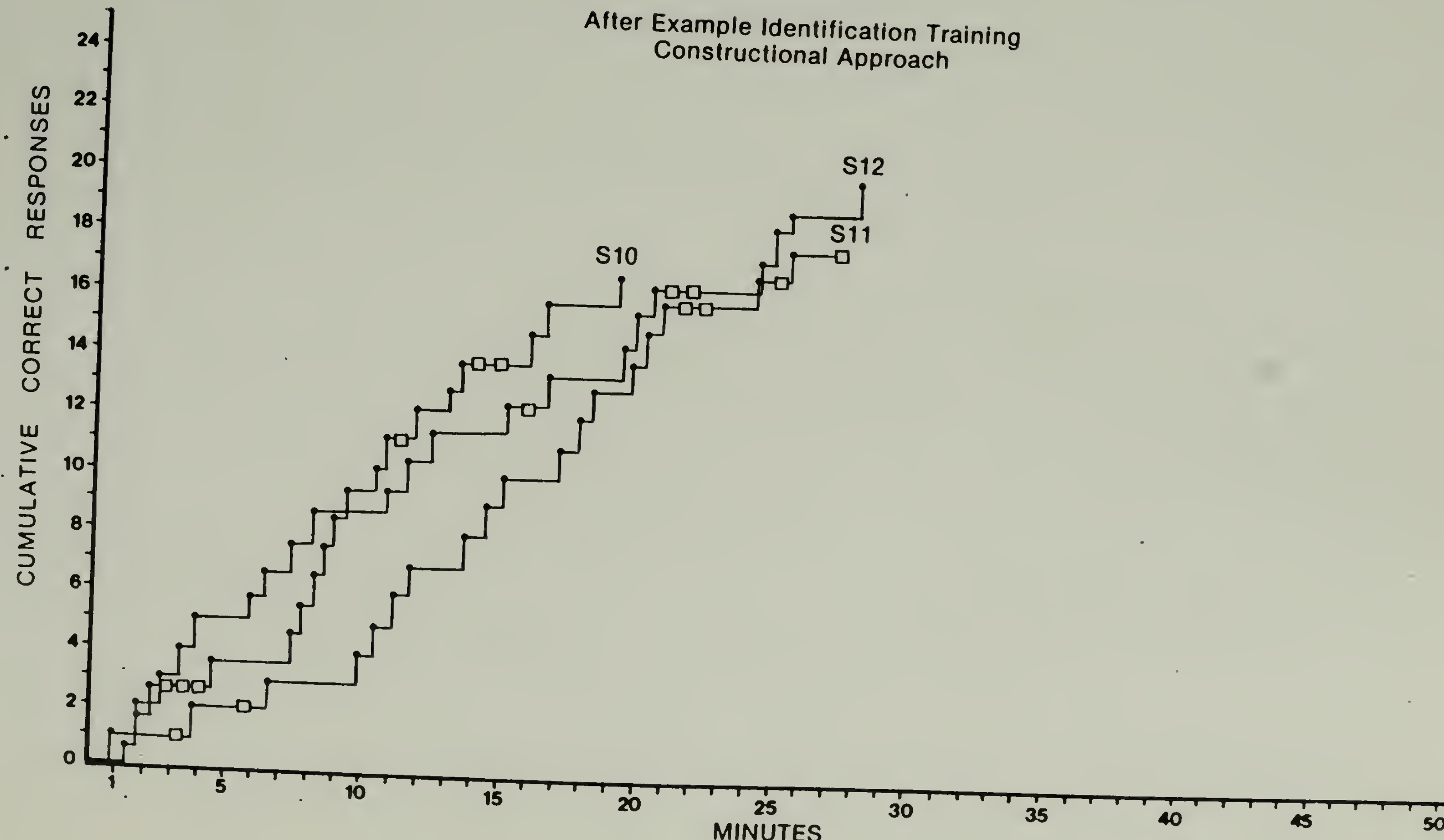


Figure 21. Cumulative frequency of correct test performance.


After Definition/Example Identification Training
Tau Effect
S10
S11
S12
CUMULATIVE CORRECT RESPONSES
0
2
4
6
8
10
12
14
16
18
20
22
24
MINUTES
1
5
10
15
20
25
30
35
40
45
50


After Example Identification Training
Constructional Approach
CUMULATIVE CORRECT RESPONSES
MINUTES
S10
S11
S12
0
2
4
6
8
10
12
14
16
18
20
22
24
1
5
10
15
20
25
30
35
40
45
50

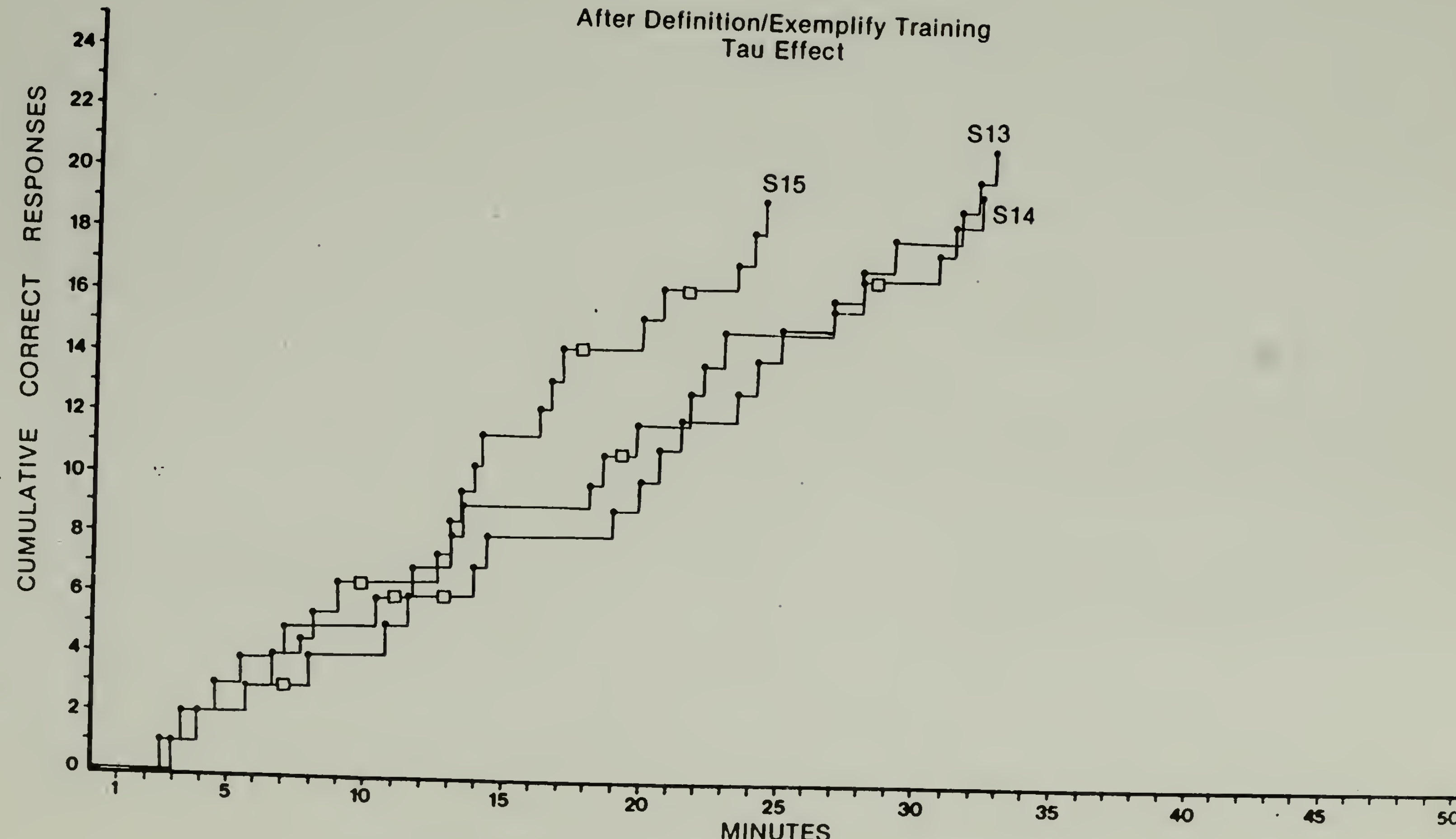
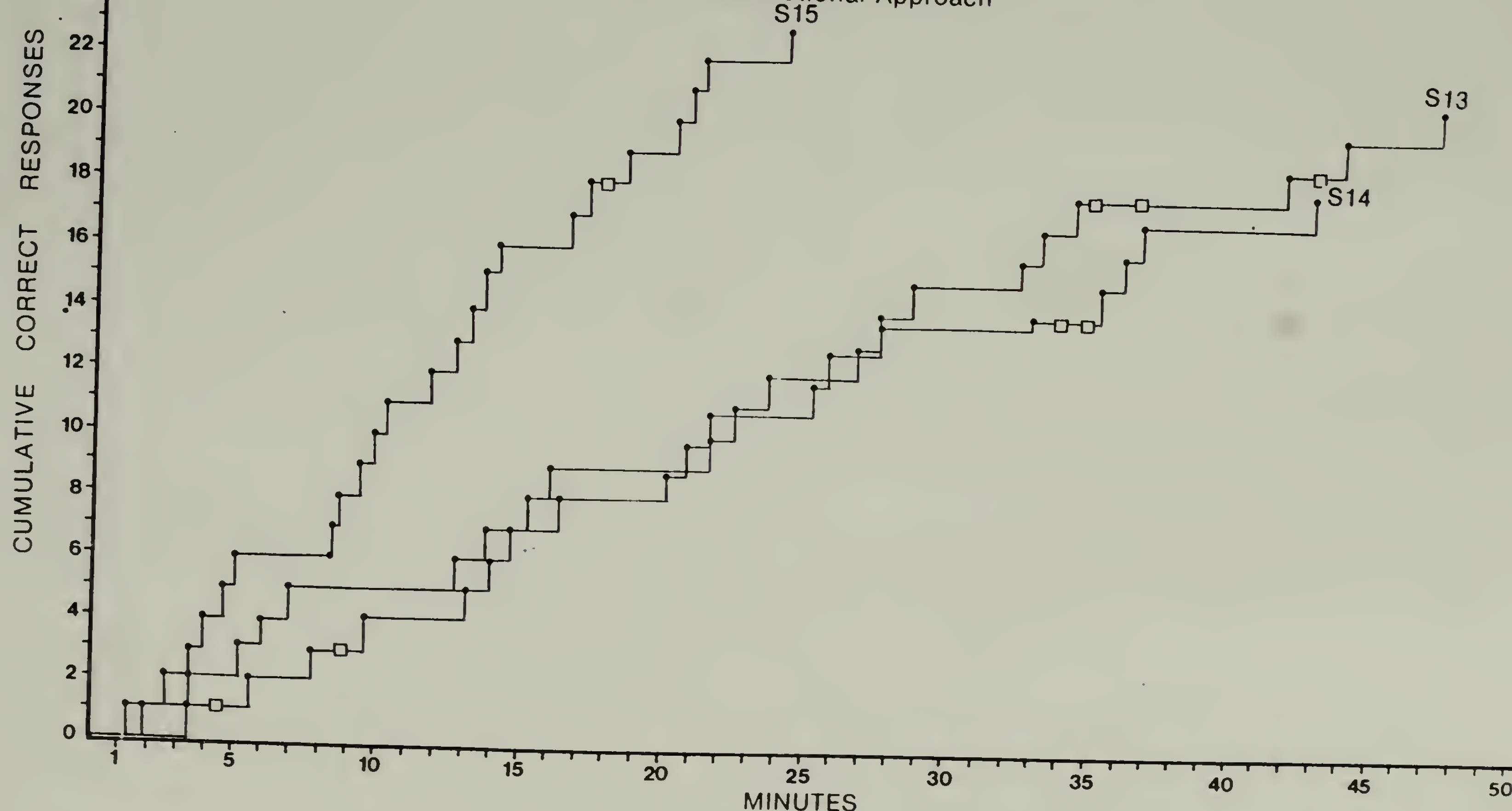
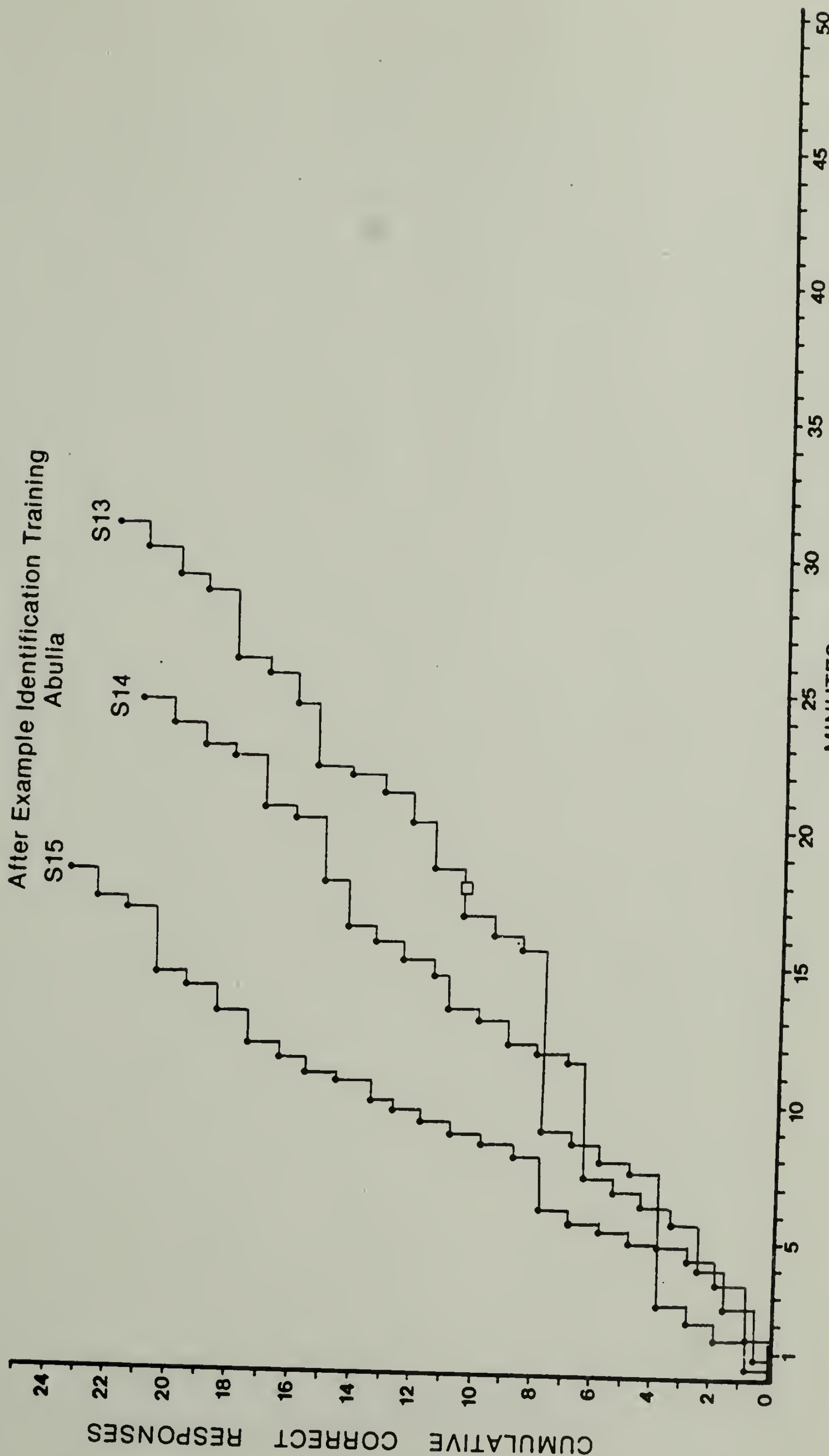


Figure 22. Cumulative frequency of correct test performance.

After Definition/Example Identification Training
Constructional Approach




After Example Identification Training
Abulia
S15
S14
S13
CUMULATIVE CORRECT RESPONSES
MINUTES
0
2
4
6
8
10
12
14
16
18
20
22
24
1
5
10
15
20
25
30
35
40
45
50

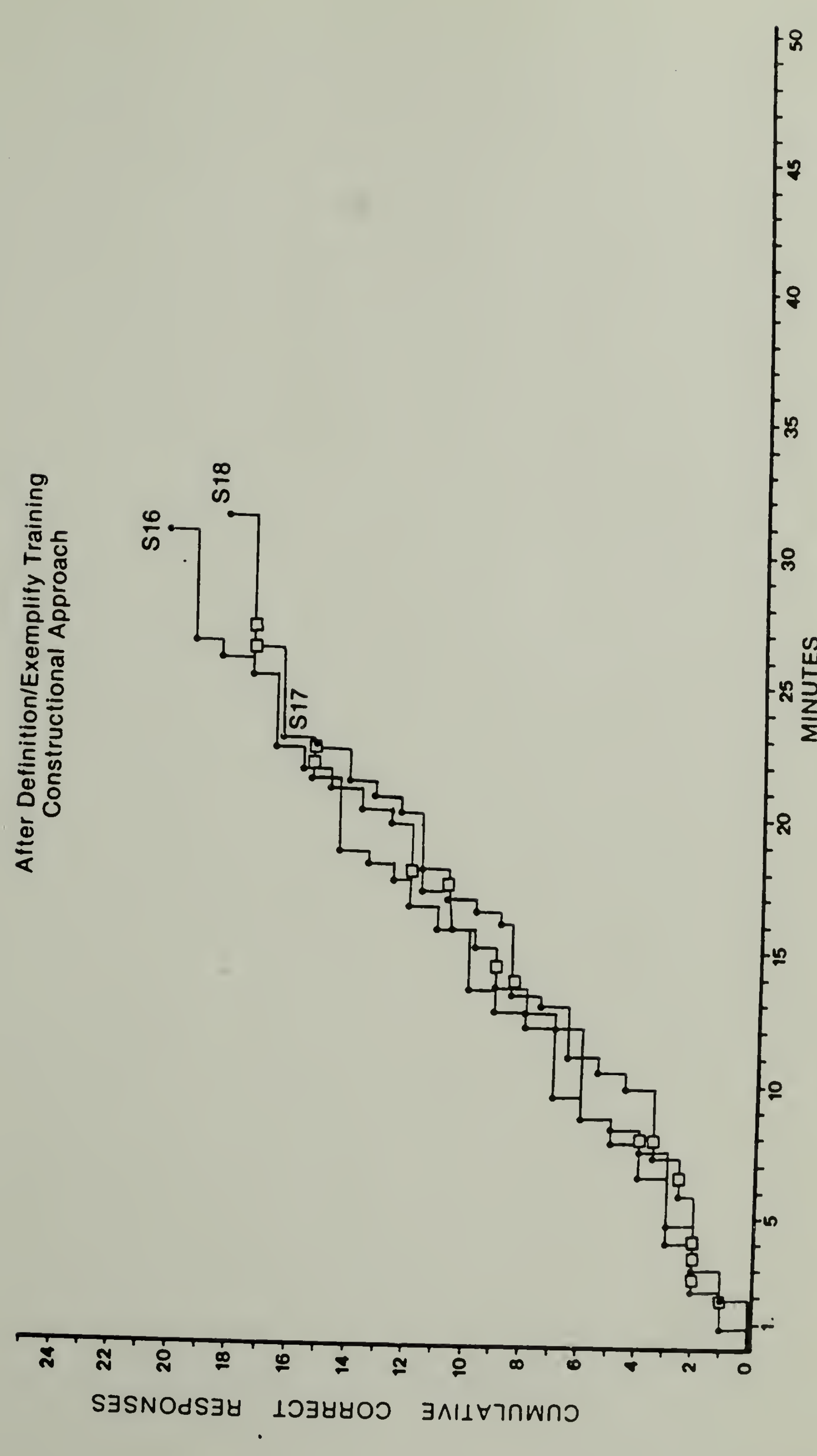
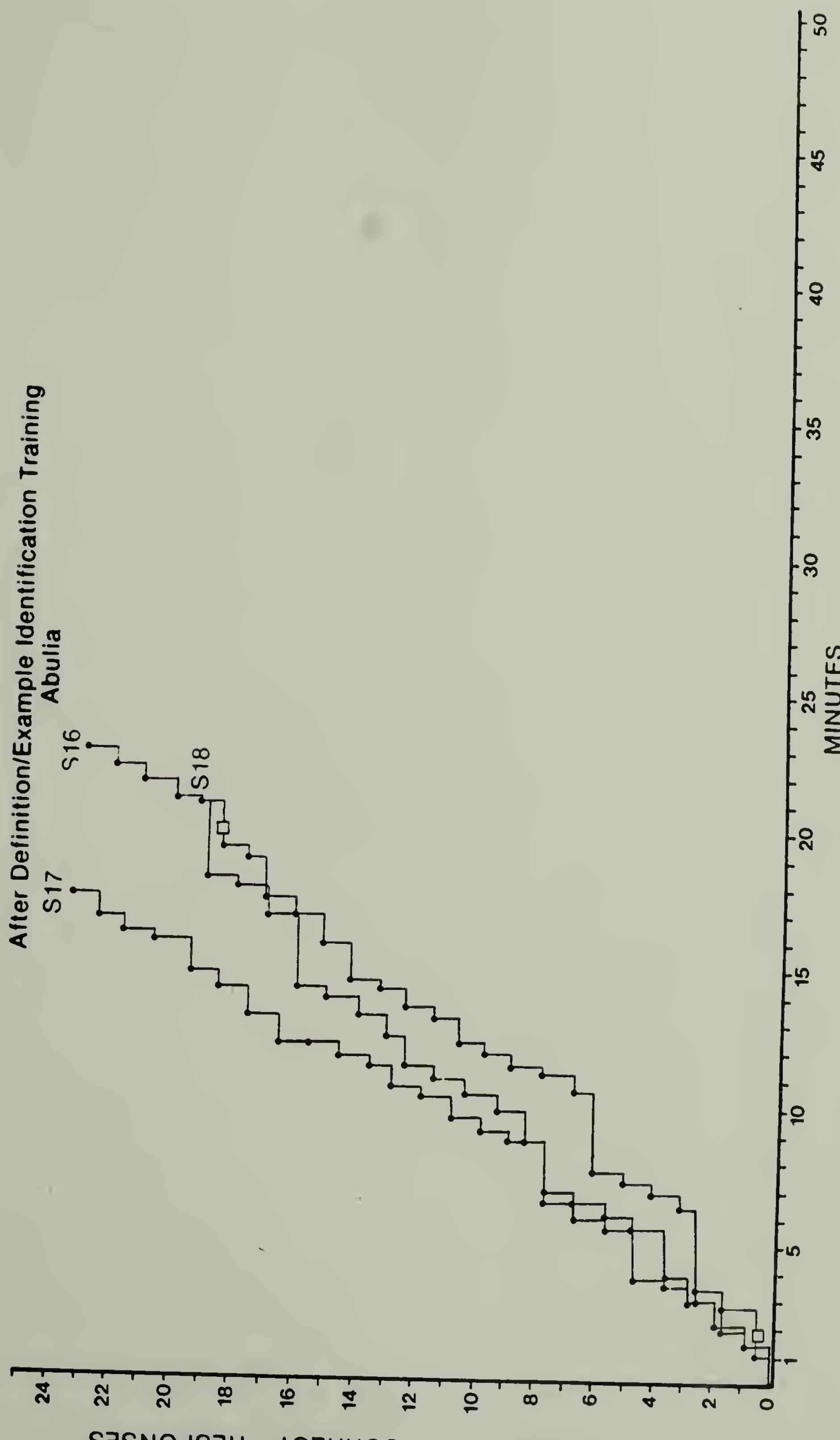
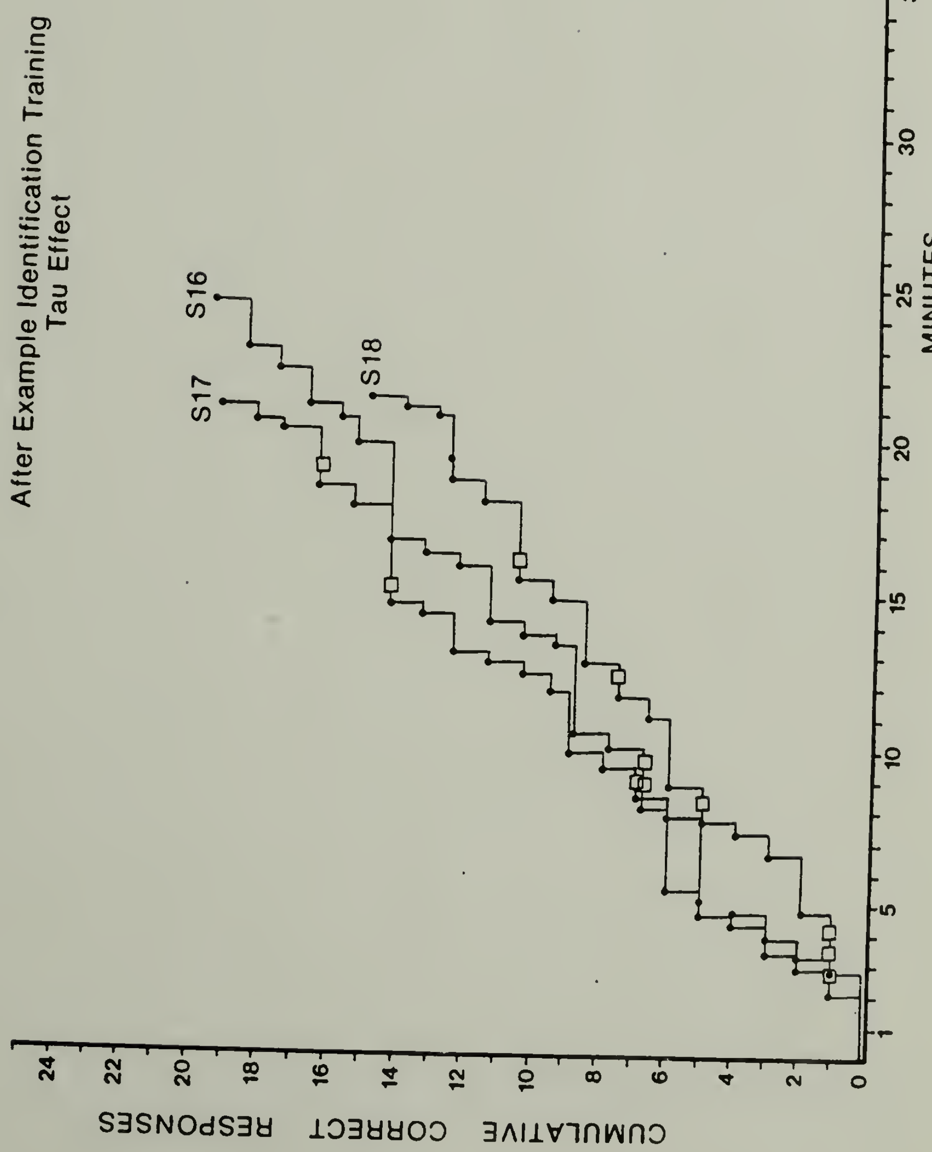


Figure 23. Cumulative frequency of correct test performance


After Definition/Example Identification Training
Abulia
S17
S16
S18
CUMULATIVE CORRECT RESPONSES
0
2
4
6
8
10
12
14
16
18
20
22
24
MINUTES
1
5
10
15
20
25
30
35
40
45
50


After Example Identification Training
Tau Effect
S16
S17
S18
CUMULATIVE CORRECT RESPONSES
0
2
4
6
8
10
12
14
16
18
20
22
24
MINUTES
1
5
10
15
20
25
30
35
40
45
50

were not found. Five subjects highest rates occurred after definition/exemplify training. Seven subjects highest rates occurred after definition/example identification training and six subjects highest rates occurred after example identification training.

Second, systematic differences were found between concepts. Seventeen of the eighteen subjects responded faster on the abulia test than on either the tests for constructional approach or tau effect. The sole exception had very similar rates for all three concepts (Subject 12). Fifteen of the eighteen subjects responded faster on tau effect tests than on constructional approach tests. Two of the three exceptions had equal rates for tau effect and constructional approach. A planned comparison between abulia and the other two concepts was significant, $F(1, 22) = 85.29$, $p < .01$. In addition, a planned comparison between tau effect and constructional approach revealed a significant difference, $F(1, 22) = 20.04$, $p < .01$.

Figures 18-23 also present the total duration of each test for the eighteen subjects. These data reveal systematic relations between concepts, but not between study programs. Fifteen subjects spent less time on the tests for abulia than on the tests for either of the other two concepts. In addition, sixteen subjects spent less time on the tests for tau effect than on the tests for constructional approach. Planned comparisons between concepts revealed a significant difference between abulia and the other two concepts, $F(1, 22) = 39.18$, $p < .01$ and a significant difference between tau effect and constructional approach, $F(1, 22) = 5.13$, $p < .05$.

Figures 24 and 25 present the percent correct on the total test for each of the eighteen subjects. Graphing conventions are the same as Figures 16 and 17. The data revealed that ten subjects' had higher percent correct scores after definition/exemplify training than after example identification training. In addition, twelve subjects had higher scores after definition/example identification training than after example identification training. In addition, no systematic differences were found between definition/exemplify training and definition/example identification training. However, planned comparisons revealed that neither of the two relations were significant.

Similar results were evident for comparisons between concepts. Seven subjects had higher scores on abulia than either of the other concepts, and seven subjects had higher scores on tau effect than either of the other two concepts. However, neither of these relations was significant.

- - - - - - - - - - - - - - - - - -

SEE FIGURES 24 and 25, PAGES 193-194

- - - - - - - - - - - - - - - - - -

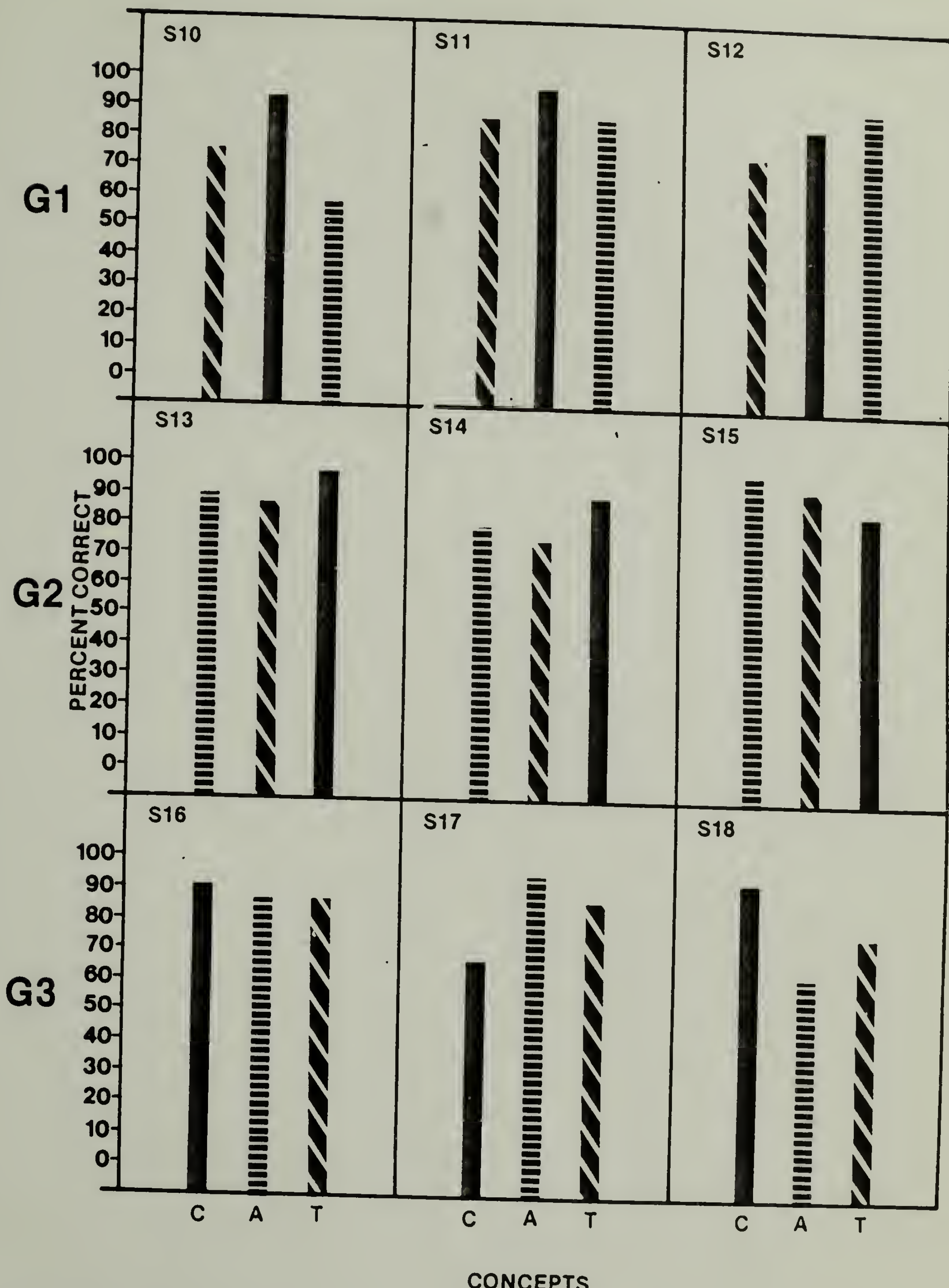


Figure 24. Total percent correct on test for subjects who did *not* receive a pretest.

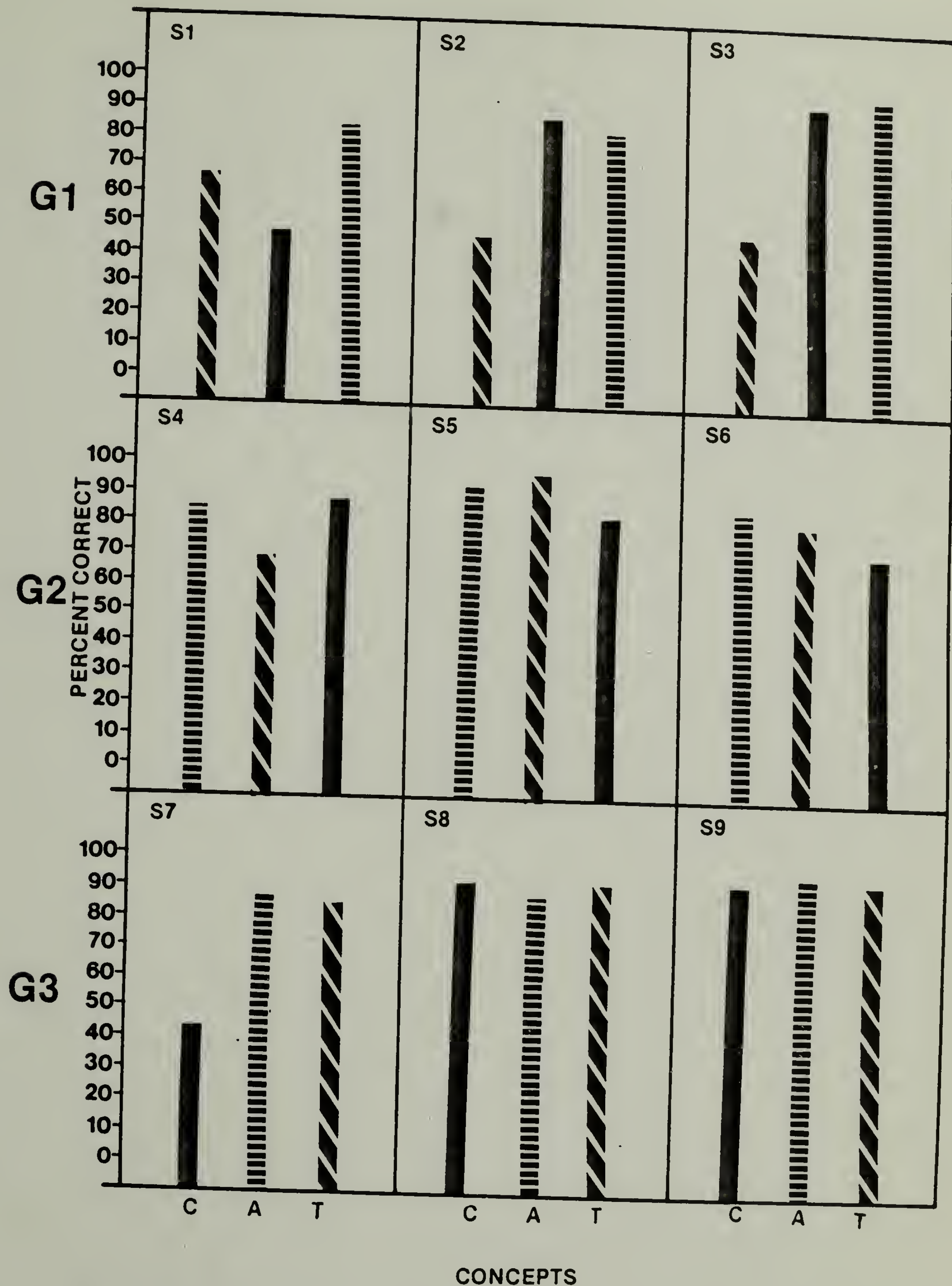


Figure 25. Total percent correct on test for subjects who did receive a pretest.

Test performance was also subjected to separate analyses of extension and transfer. Extension was defined as performance on the types of tasks for which training was received. Transfer was defined as performance on combination tasks. Extension performance was analyzed to determine whether the differences found during training were maintained under test conditions. Transfer was analyzed to determine the differential effects of the programs on performance of a class of tasks that was not explicitly taught. As usual, both rate of correct responding and percent correct measures were analyzed.

Figures 26 and 27 present the correct responses per minute on extension tasks for all eighteen subjects. The same graphing conventions are used here as were used previously. Figure 26 presents data for subjects who received a pre-test, Figure 27 presents data for subjects who did not. These data reveal both systematic differences between study programs and between concepts.

- - - - - - - - - - - - - - - - - - -

SEE FIGURES 26 AND 27, PAGES 196-197

- - - - - - - - - - - - - - - - - - -

First, eighteen subjects had lower rates on tests after definition/exemplify training than on tests after example identification and definition/example identification training. Twelve subjects had higher rates after example

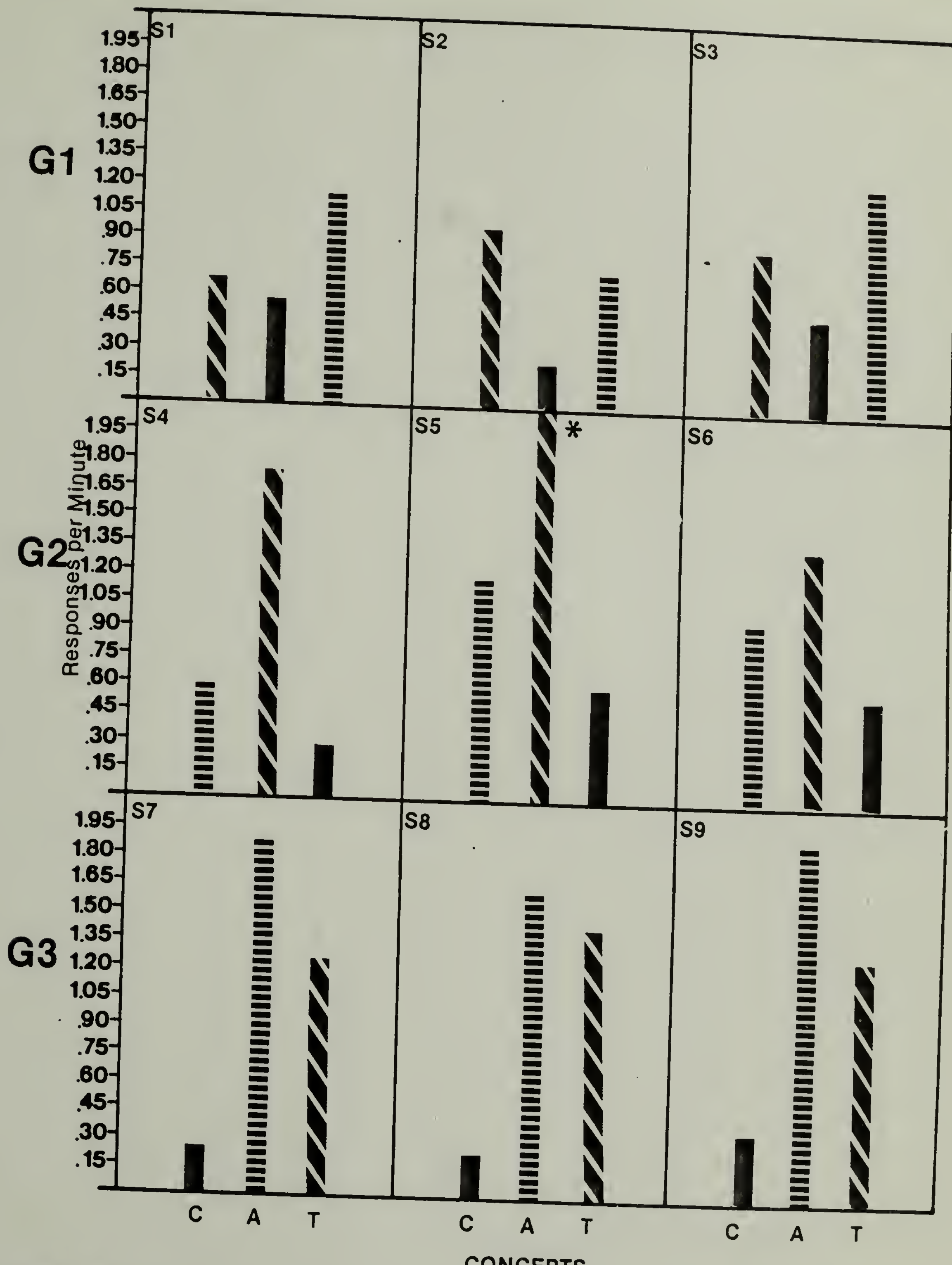


Figure 26. Correct responses per minute on extension tasks for subjects who received a pretest.

*Greater than 2.0 responses per minute.

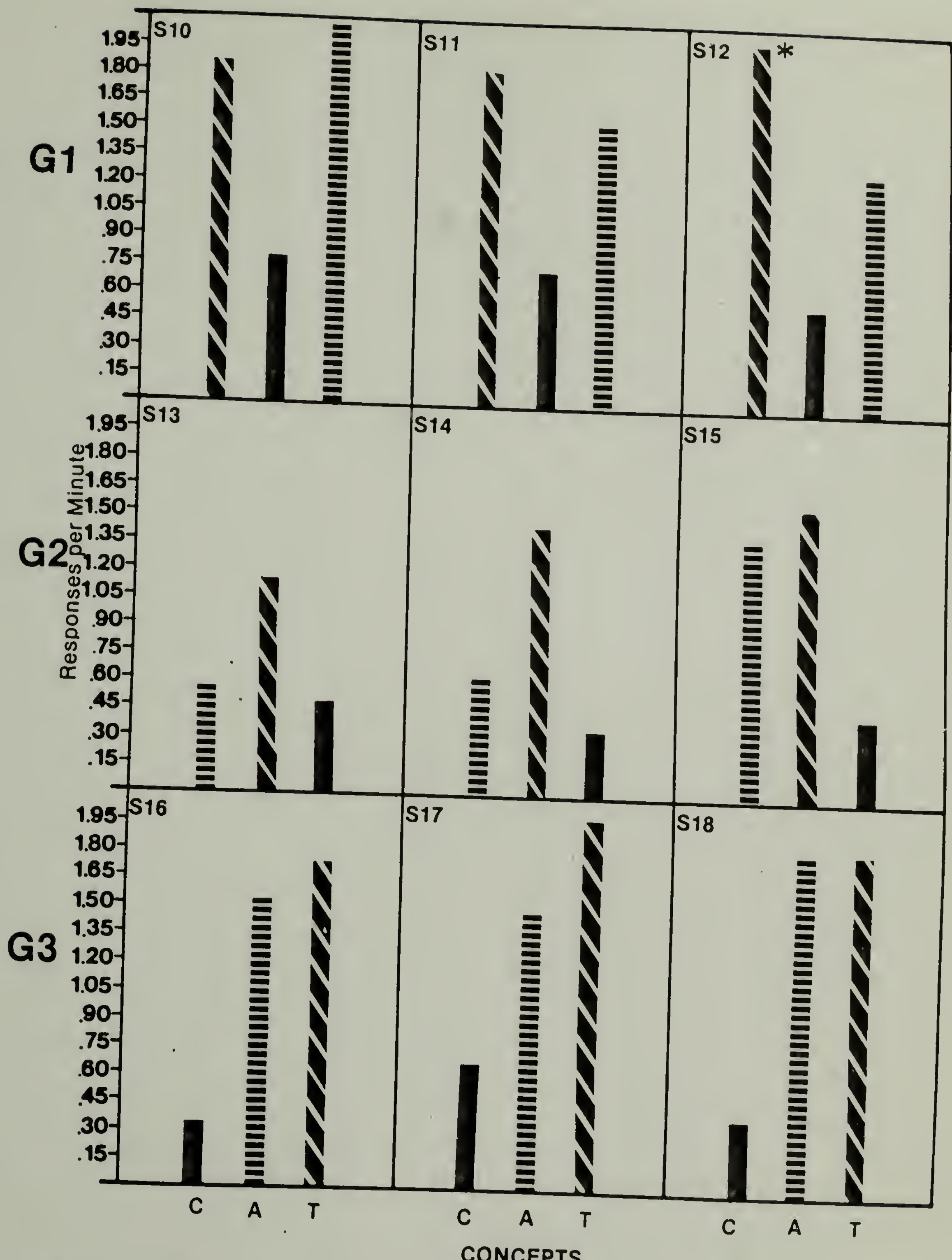


Figure 27. Correct responses per minute on extension tasks for subjects who did *not* receive a pretest.

*Greater than 2.0 responses per minute.

identification training than after definition/example identification training. Three of the six exceptions (Subjects 1, 3, and 10) occurred when example identification training, occurred with constructional approach. One of the remaining exceptions (Subject 9) had low study performance (57% correct) and a moderate rate of correct responding. (.70 per minute) on study tasks for the example identification program. Planned comparisons between study programs revealed a significant difference between definition/exemplify training and both of the other programs, $F\ (1, 22) = 220.84$, $p < .01$, and a significant difference between example identification training and definition/example identification training, $F\ (1, 22) = 8.80$, $p < .01$.

Second, twelve subjects rates of correct responding were higher on test items for abulia than on <u>constructional approach</u>. All six exceptions (Subjects 1-3; 10-12) occurred when abulia was combined with definition/exemplify training. Nine subjects rates were higher on <u>tau effect</u> tests than on <u>constructional approach</u> tests, yet six of these occurred when constructional approach was combined with definition/exemplify training. Regardless, a planned comparison between <u>constructional approach</u> and both <u>abulia</u> and <u>tau effect</u> revealed a significant difference, $F\ (1, 22) = 25.18$, $p < .01$. No difference was found between <u>tau effect</u> and <u>abulia</u>.

Figures 28 and 29 present the percent correct extension performance for all eighteen subjects. Graphing conventions are the same as those described for the other figures. No systematic differences were found between study programs. However, there were differences between concepts.

Specifically, thirteen subjects had lower percent correct performance on <u>constructional approach</u> than either <u>abulia</u> or <u>tau effect</u>. Two of the exceptions to this relation (Subjects 1 and 15) had very similar performance on all three concepts. The remaining three exceptions (Subjects 16-18) all had a combination of <u>constructional approach</u> and definition /exemplify program and none of the three received a pretest. A planned comparison between constructional approach and the other concepts revealed a significant difference, $F\ (1,\ 22) = 14.45$, $p < .01$.

- - - - - - - - - - - - - - - - - - - -

SEE FIGURES 28 AND 29, PAGES 200-201

- - - - - - - - - - - - - - - - - - - -

Figures 30 and 31 present the correct responses per minute data for combination tasks. Graphing conventions are the same. These data revealed no systematic differences between concepts or study programs. However, they do reveal transfer performance that is much lower than corresponding extension performance. On fifty of the fifty-four tests, extension performance was higher than transfer performance. A planned comparison between extension and transfer scores revealed significantly higher extension scores, $t\ (106) =$

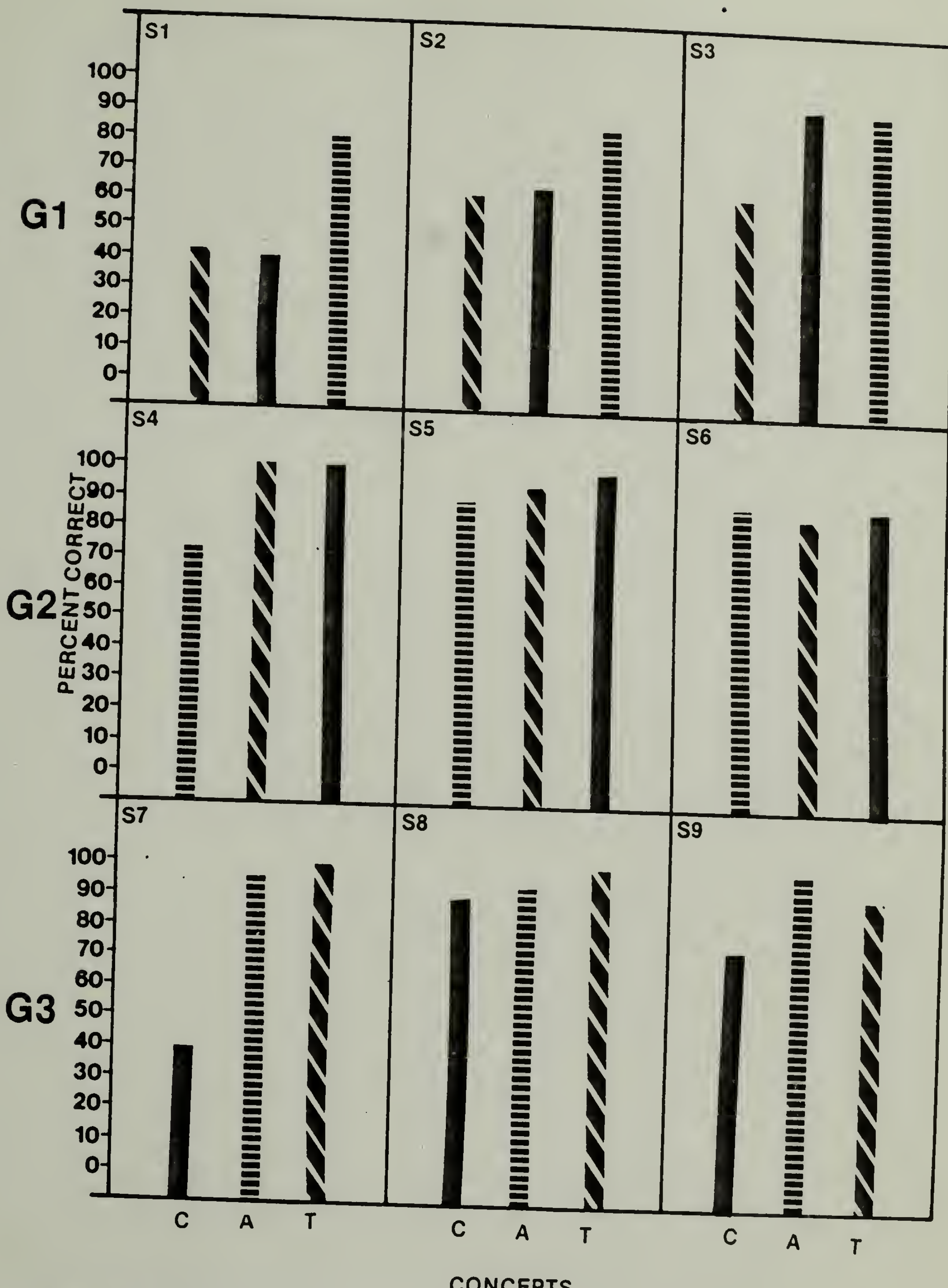


Figure 28. Percent correct on extension tasks for subjects who received a pretest.

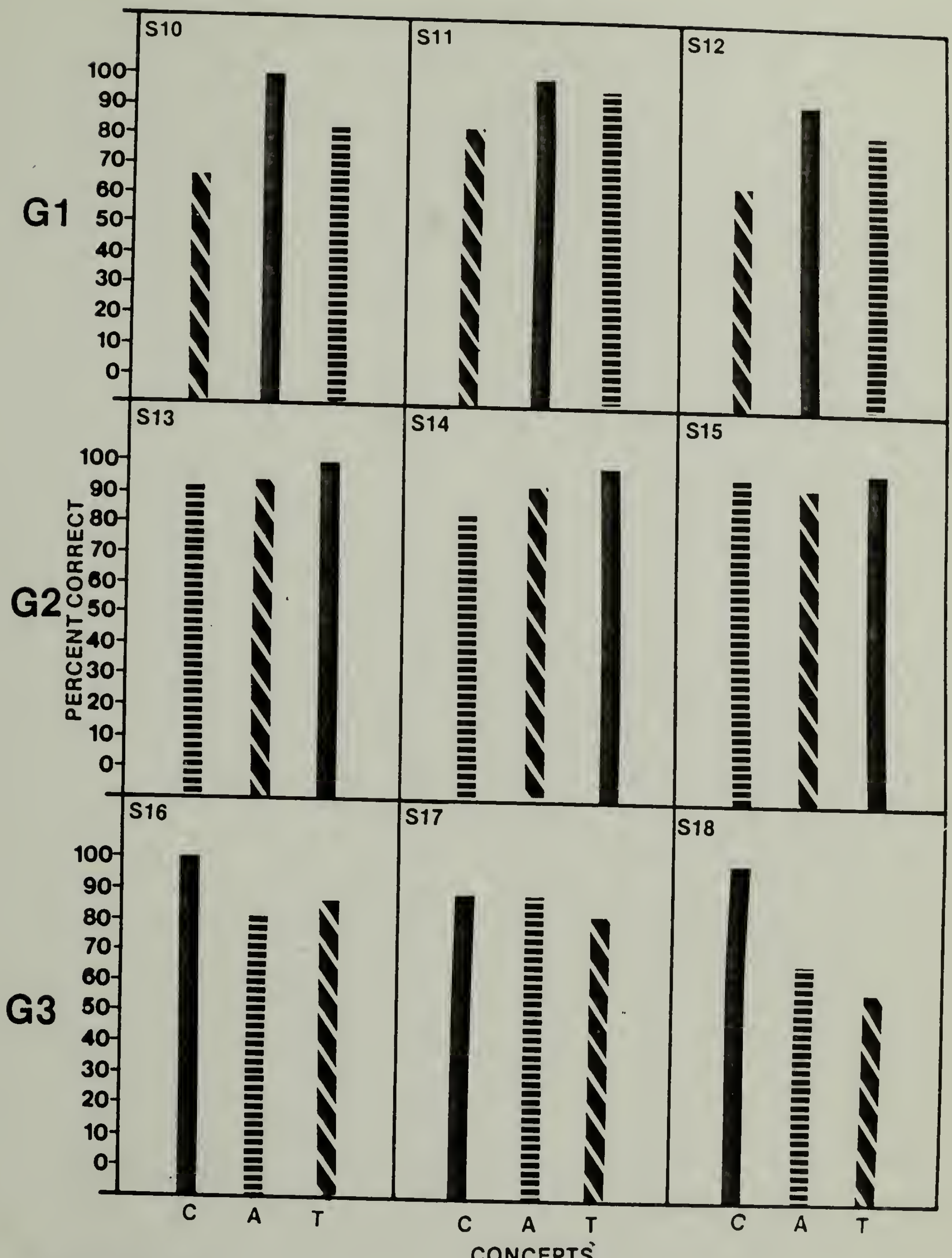


Figure 29. Percent corect on extension tasks for subjects who did *not* receive a pretest.

8.05, $p < .001$.

- - - - - - - - - - - - - - - - - - -

SEE FIGURES 30 AND 31, PAGES 203-204

- - - - - - - - - - - - - - - - - - -

Figures 32 and 33 present the percent correct performance on combination tasks. Graphing conventions are the same. As with the rate data, these data revealed no systematic differences between concepts or study programs and most subjects performed better on extension items than they did on combination items. However, this latter finding was not as systematic as the rate data. Only 37 of the fifty-four tests had higher extension performance than transfer performance. Of the 17 exceptions, though, three had equal performance, (Subjects 8, 13 and 18) and six had above 90% performance on both types of items (Subjects 5, 11, 12, 15 and 17). Regardless, a planned comparison between extension and transfer scores revealed significantly higher extension performance, $t(106) = 4.19$, $p < .001$.

- - - - - - - - - - - - - - - - - - -

SEE FIGURES 32 AND 33, PAGES 205-206

- - - - - - - - - - - - - - - - - - -

Further group analyses were conducted to statistically substantiate the comparisons described above. A four way, repeated measure, latin square ANOVA was conducted for each dependent measure. Specifically, seven 2 (pretest) x 3

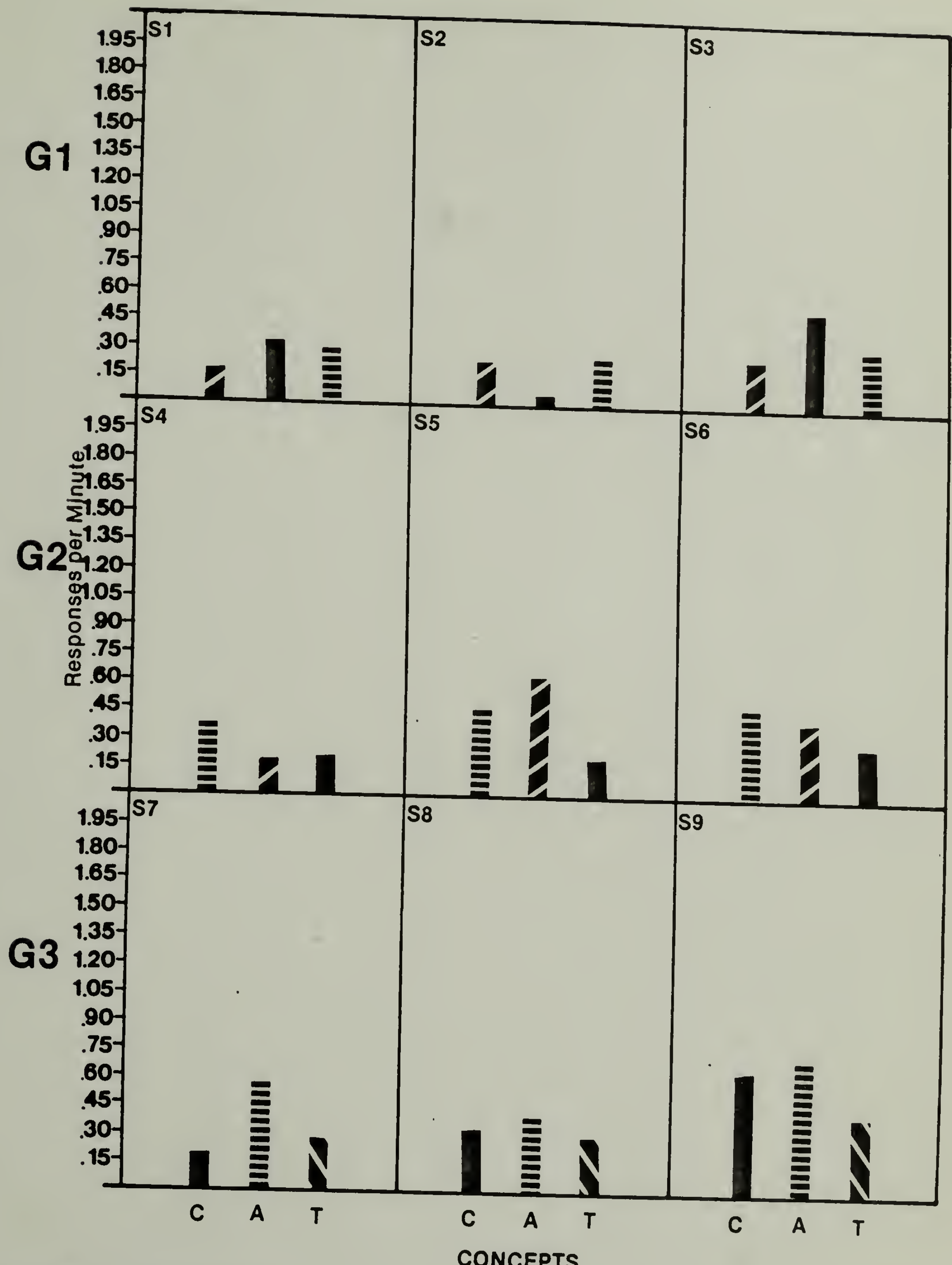


Figure 30. Correct responses per minute on combination tasks for subjects who received a pretest.

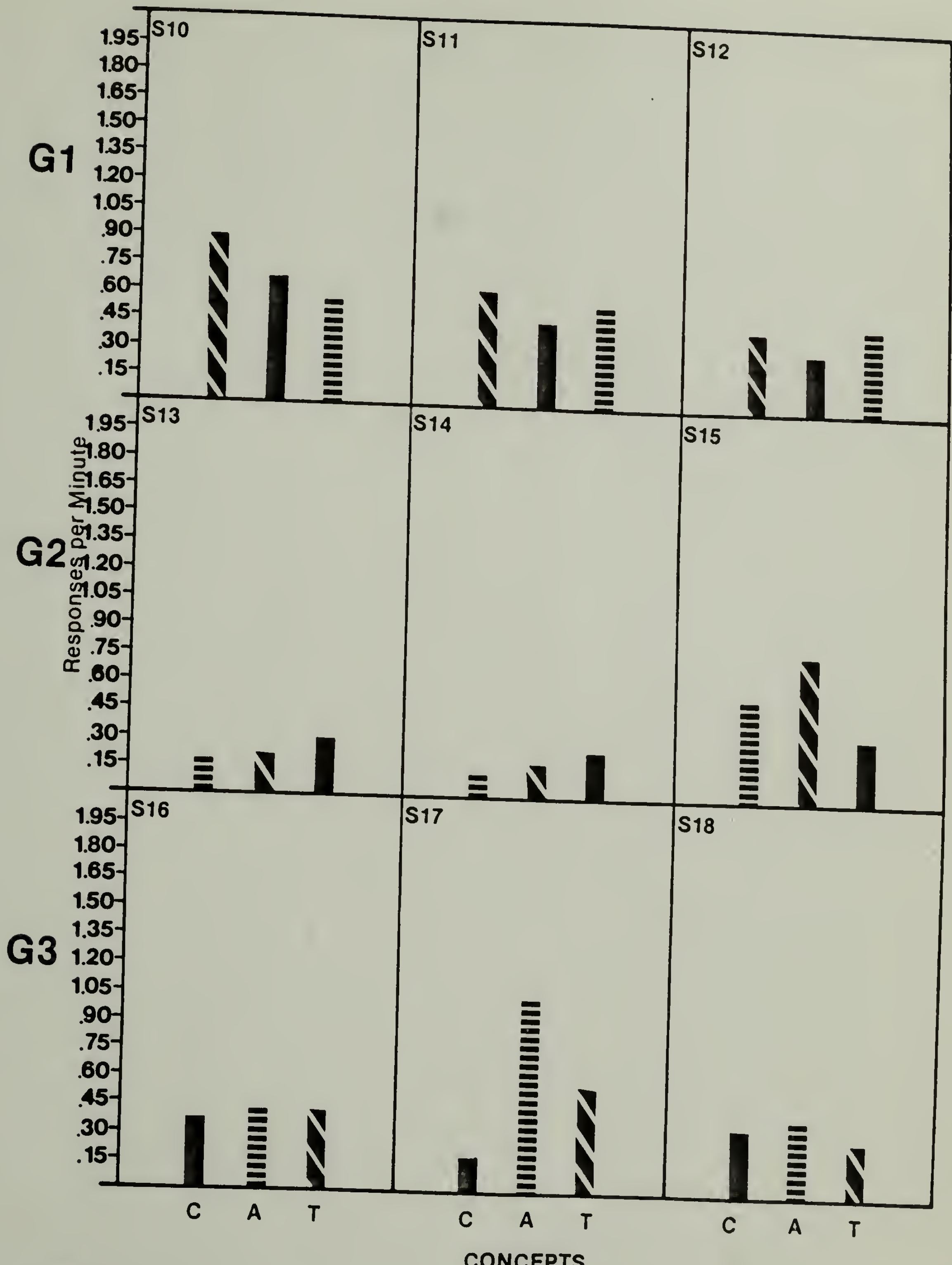


Figure 31. Correct responses per minute on combination tasks for subjects who did *not* receive a pretest.

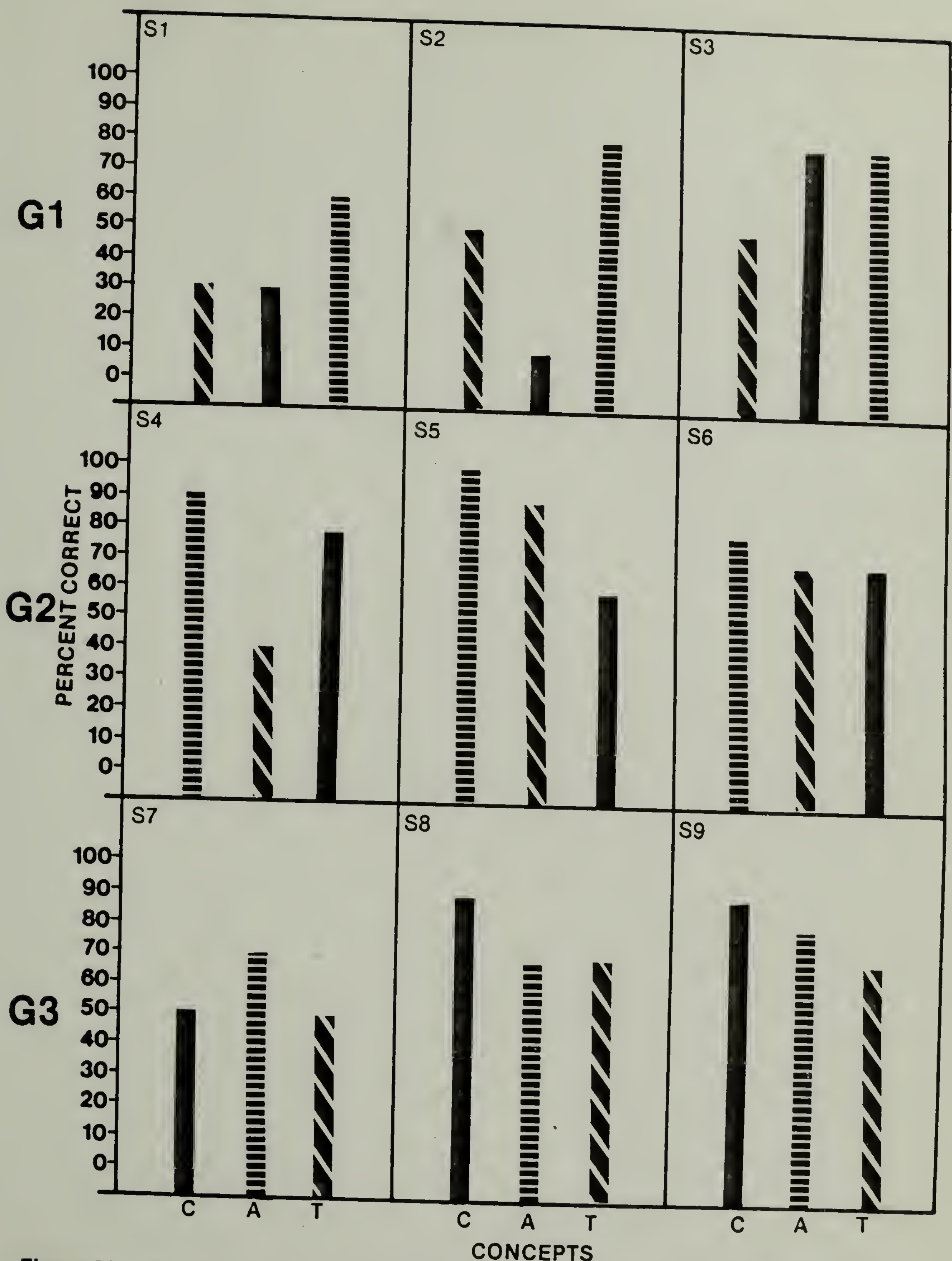


Figure 32. Percent correct on combination tasks for subjects who did receive a pretest.

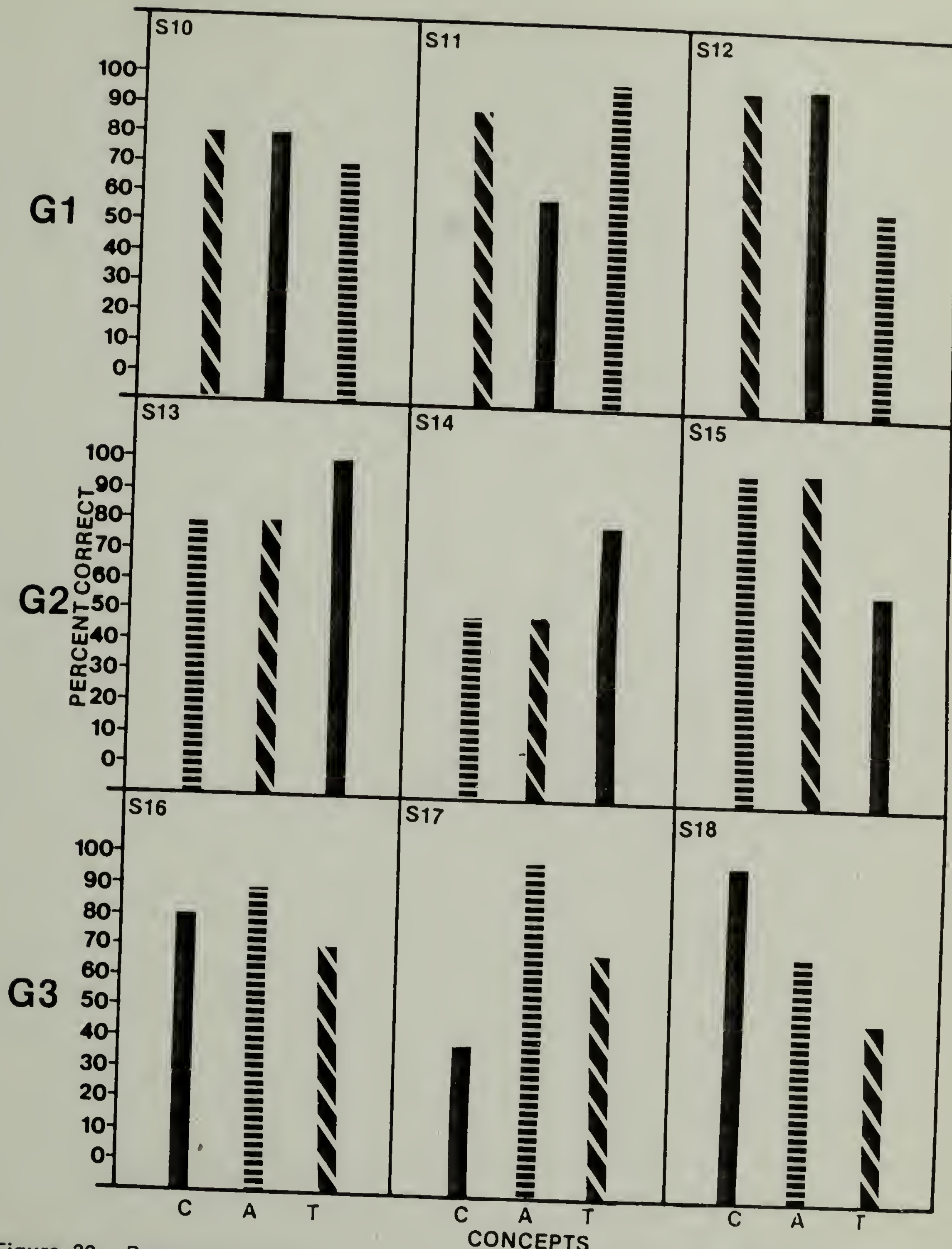


Figure 33. Percent correct on combination tasks for subjects who did *not* receive a pretest.

order of study program presentation) x 3 (concepts) x 3 (study program) ANOVA's were calculated. An arc sin transformation of proportions was calculated for all percent correct data to stabilize the variance in order to assume homogeneity of variance. The different analyses revealed different significant effects.

First, the ANOVA of correct responses per minute for the total test revealed a significant concept difference, $F(2, 22) = 50.82$, $p < .01$. However, no significant effects were found for pretest or order of study program and no significant interactions were found. As described earlier, a planned comparison between concepts demonstrated a significant difference between _abulia_ and the other concepts and a significant difference between _tau effect_ and _constructional approach_.

Second, the ANOVA of percent correct performance for the total test revealed a significant study program by pretest interaction, $F(2, 22) = 5.18$, $p < .05$. Apparently, subjects who did not receive the pretest scored significantly higher on tests after example identification training than subjects who received the pretest. No other significant effects or interactions were found for the total test percent correct data.

Third, the ANOVA for duration of total test revealed a significant difference between concepts, $F(2, 22) = 22.15$, p $<.01$. As described above, planned comparisons of duration measures between concepts revealed a significant difference between abulia and the other two concepts and a significant difference between tau effect and constructional approach. No other significant effects or interactions were found for the duration data.

Fourth, the rate of correct responding for extension items revealed a number of significant effects and interactions. Table 20 presents the source data for this analysis. It can be seen that significant study program, concept and pretest differences were found, $F(2, 22) = 114.80$, $p < .01$, $F(2, 22) = 15.29$, $p < .01$ and $F(2, 10) = 7.07$, $p < .025$ respectively. Planned comparisons between study programs and between concepts were described above. In addition, significant order by pretest and concept by pretest interactions were revealed, and an order by concept interaction was estimated. The possible interaction between concepts and order estimated by the BC residual effect and the interaction between pretest and order, prohibit definitive conclusions about the main effects since the F ratios for the main effects are positively biased when such interactions occur.

- - - - - - - - - - - - - - -

SEE TABLE 20, PAGE 209

- - - - - - - - - - - - - - -

TABLE 20

ANALYSIS OF VARIANCE FOR RATE OF CORRECT RESPONDING ON EXTENSION TASKS

Source	Sum of Squares	DF	Mean Squares	F
Mean	65.340	1	65.340	697.84
Pretest (A)	.662	1	.662	7.07**
Order (B)	.469	2	.235	2.51
AB	1.682	2	.941	8.98***
Error	1.123	12	.094	
Concept (C)	1.1564	2	.782	15.29***
AC	.781	2	.390	7.63***
BC				
Program (D)	11.480	2	5.740	114.80***
Residual	.257	2	.129	2.57
ABC				
AD	.180	2	.09	1.80
Residual	.055	2	.027	.55
Error	1.227	24	.051	

** indicates $p < .025$

*** indicates $p < .01$

Fifth, the percent correct data for extension items revealed a number of significant effects and interactions. Table 21 presents the source data for this analysis. It can be seen that significant concept and order effects were found, $F\ (2,\ 22) = 7.36$, $p < .01$ and $F\ (2,\ 10) = 5.64$, $p < .025$ respectively. Planned comparisons between concepts have been described already. Planned comparisons between different orders of study program presentation revealed significant lower performance for order I (definition/exemplify, definition/example identification, example identification) than the other two orders $F\ (1,\ 10) = 8.64$, $p < .025$. No difference was found between orders II and III. In addition, significant interactions were found between concept and pretest, $F\ (2,\ 22) = 5.31$, $p < .025$, and pretest and order of study program. Both the main effect of order and the interaction between order and pretest indicate that the F ratios for the main effects in this analysis are inflated.

- - - - - - - - - - - - - - - -

SEE TABLE 21, PAGE 211

- - - - - - - - - - - - - - - -

Sixth, both rate of correct responding and percent correct performance on combination tasks were also analyzed. However, no main effects or interaction effects were found for either of these sets of data.

TABLE 21

ANALYSIS OF VARIANCE FOR PERCENT CORRECT PERFORMANCE ON EXTENSION ITEMS

Source	Sum of Squares	DF	Mean Squares	F
Mean	396465.352	1	296465.352	2148.22
Pretest (A)	411.129	1	411.129	2.23
Order (B)	2086.037	2	1043.019	5.65**
AB	851.815	2	425.907	2.31
Error	2214.667	12	184.556	
Concept (C)	1671.370	2	835.685	7.36***
AC	1205.815	2	602.907	5.31**
BC				
Program	431.13	2	215.560	1.89
Residual	306.27	2	153.14	1.35
ABC				
AD	1895.50	2	947.80	8.35***
Residual	73.90	2	36.95	.33
Error	2724.000	24	113.50	

** indicates $p < .025$

*** indicates $p < .01$

In summary, both group and intrasubject analyses of test data revealed aystematic differences in performance. First, rate of correct responding on the total test, duration of the total test, rate of correct extension performance all revealed the same differences between concepts. Most subjects were slower and less accurate with items on constructional approach than they were on both abulia and tau effect items. In addition, subjects had lower rates of correct performance and took longer to complete tau effect items than abulia items. Second, most subjects' rate of correct extension performance was faster after example identification training than after definition/exemplify training or definition/example identification training. In addition, subjects were faster on definition/example identification extension tasks than on definition/exemplify extension tasks. Third, both rate of correct extension performance and percent correct extension performance were higher than corresponding transfer performance. Fourth, those subjects who received the pretest had lower rates of extension performance than those who did not receive the pretest. Fifth, order of study program presentation appeared to effect percent correct extension performance. Subjects who received order I (definition/exemplify, definition/example identification, example identification) performed lower than subjects who had received either of the other two orders. Finally, a number of interactions were found.

Interactions between pretest and order and between pretest and concepts were found on rates of correct extension performance and percent correct extension performance. In addition, an interaction between study program and pretest was found on percent correct measures of total test performance.

Discussion

Experiment 2 was conducted for two reasons: 1) to compare acquisition and test performance under three different study programs and 2) to refine the methodology for testing the effects of study programs. It appears that some definitive conclusions can be discussed for each. Comparisons among study performances and among test performances revealed systematic differences while demonstrating control over many extraneous variables. However, the results do suggest further refinements, modifications and additional questions that should also be discussed. Therefore, this section is structured in order to emphasize three kinds of conclusions. First, the comparisons among study programs and between extension and transfer performance are described. Second, methdological issues are discussed and finally, future studies and modifications of these investigations are suggested.

Comparisons among study programs. The most important contributions of these data that are the differences among types of questions that were found in experiment 1 were reproduced and extended. These findings can be categorized simply. First, example identification tasks were more efficient for the student than other tasks: both the study programs that contained example identification tasks and the example identification extension items took less time to correctly complete than other types of questions. Second, both definition and exemplify questions were answered more accurately than example identification questions. Subjects made fewer mistakes on these types of tasks than on example identification questions during the study trials and this accuracy was maintained on the tests. Therefore, the most accurate performances occurred when definition and exemplify questions were combined in a study program. Third, differences in overall test performance and performance on transfer items were not systematically related to the type of training that subjects received. The within subject differences that existed across sessions were probably attributable to other factors such as concept, whether or not the subject received a pretest and the order in which the study programs were received. Other factors that might be generally categorized as historical factors or entering skills may also have contributed to the different effects that study programs had across subjects.

These relations indicated that the same possible generalizations specified in experiment 1 can be extended to comparisons of example identification programs with two types of multiple question programs. When efficiency or cost-effectiveness for the student is an important factor for the design of study materials, then example identification programs are better than either definition/exemplify or definition/example identification programs. However, in cases for which accuracy of study performance is a more important criterion, then both definition and exemplify questions would be the study questions of choice. If both efficiency and accuracy are critical, combining definition and example identification questions within a study program will achieve a middle ground. Obviously, these generalizations need to be explicitely tested in applied settings.

As discussed in experiment 1, it is important to emphasize the difference between student efficiency and teacher efficiency. The data indicated the efficiency of example identification programming for students. However, example identification programs took longer to design than any of the other programs that have been tested. In fact, the time that it took to write an extensive pool of example identification items, to test them, to rewrite some and to select a range of easy to difficult items may be prohibitive for some teachers. Therefore, if example identification programs are to be used, alternative strategies are needed for developing them.

One strategy that might be acceptable to some teachers is to enlist their students' help in writing example identification questions. Since definition/exemplify study programs do facilitate total test and transfer performance as well as example identification programs, they could be used with the first group of students who come into contact with course materials. Then, as each section of the study program is completed by the students, their answers to the study questions could be collected and the exemplify answers would produce a pool of example identification questions that could be used in future programs. Of course, difficult material could still be completely programmed with example identification items. In addition, the illustrations by students generated would need to be edited and tested. Regardless, using student illustrations would minimize the time that it takes to write a pool of examples and nonexamples and perhaps would produce a wider range of illustrations.

Extension and transfer performance. Before the practical implications stated above can be seriously considered, the difference between extension and transfer performance that was found in both experiments must be scrutinized. Both results suggested that there was a systematic and significant difference between extension and transfer performance. Specifically, subjects responded faster and more accurately on extension items than on transfer items. Experiment 2 extended the findings of experiment 1 by explicitely measuring the degree of transfer to the most integrated and probably the most difficult class of task from the typology: the combination task. The data indicated that transfer from all study programs was low. Therefore, these results suggest that teachers need to considered carefully the kinds of preparations that they require of students if they want them to master such integrated tasks. Certainly, the kinds of tasks programmed and tested here were not sufficient to obtain high levels of performance on combination tasks. The question begged by this result is: what kinds of study programs do facilitate performance on combination questions? Is direct programming on combination questions necessary? The most obvious programs to test are those that contain combination tasks.

Methodological issues. During Experiment 1 the large number of methodological problems limited the conclusions that could be drawn from the data. The present study tried to rectify these problems. Some of the attempts were successful and some were not. In addition, some new issues surfaced that should be addressed. The following details the controls that were successfully implemented, those that were not and the new problems that these results suggested about the methodology.

Two problems remedied by the present study were: 1) the task confound present in transfer measures during the first experiment and 2) the ambiguity caused by low extension scores. Elimination of each of these sources of variability was necessary in order to answer the original question of this research: which study program facilitates transfer to other types of questions.

The task confound present in transfer measures was eliminated by defining transfer as performance on combination tasks. The measurement of transfer under each condition was therefore, the same: the same questions were used and these questions were selected from a class of tasks that had not been previously used in training. Thus, there were fewer competing hypotheses for lack of systematic effects of the three study programs on transfer performance. Factors other than the types of study programs tested must be responsible for the individual differences in transfer performance. In addition, the overall low transfer performance indicates that other types of study programs need to be tested in order to facilitate performance on combination tasks.

These results were further supported by the relatively high extension performances that were found in this study. High extension scores indicated that the subjects did in fact learn the task types for which they received training. Therefore, the question, "Does learning one or two kinds of tasks facilitate performance on other tasks (i.e. combination tasks)" was addressed.

Three procedural changes may be responsible for the increments in extension performance found in experiment 2. First, some study tasks were rewritten and therefore may have helped the subjects to learn the programs better. Second, a monetary incentive system was built into both the study trials and the test trials. Finally, subjects were required to have more correct answers during the study sequence. Although a formal mastery criterion was not implemented, more tasks were programmed so that if subjects made a mistake, they always had the opportunity to remediate the mistake by answering another similar question correctly. All three changes made it more likely that subjects would perform more accurately during study trials. Although the relation between extension performance and these changes was not tested, the higher extension scores do allow a better interpretation of the transfer and total test scores. The lack of systematic differences among these test scores can not be attributed to subjects not performing correctly on the tasks that were programmed.

A third methodological problem raised in Experiment 1 was that the effect of concepts could not be separated from the effect of the order in which subjects received the concept. As stated in experiment 1, previous research suggested that position effects for these concepts did <u>not</u> exist (Chase, 1980). In the present study, position and concept effects were again confounded. However, this study lends further support to a concept effect and not a position effect since the order of concept presentation was changed across experiments and the same concept differences were found. Specifically, in experiment 1 constructional approach was presented as the third concept and found to be more difficult than abulia or tau effect. In experiment 2 constructional approach was presented as the first concept and still found to be more difficult.

This difference in concepts remains as the major methodological problem of these investigations. To reiterate, concept differences need to be controlled if individual subject data are to be analyzed. When significant concept differences are found across subjects, then the effect of study programs can not be separated from the effect of the concept for a single subject. Therefore, intrasubject analyses are ambiguous. Because such concept effects were found in both experiments 1 and 2, the effects of study programs have to be analyzed through group, statistical procedures. Such group analyses limit the generalizations that can be made from these data to other individuals within the same population. In addition, the fact that group analyses were needed suggested that other variables and other designs should be investigated.

The variable that most obviously needs to be investigated further is the effect of different concepts. So far, the concepts used in these investigations have been tested and analyzed at least four times. First, the three concepts were chosen from a pool of concepts because subjects answered exemplify tasks with similar accuracy (Chase, 1980). Second, the three concepts were tested for similarity by having a group of subjects respond to example identification questions (Chase, 1980). Then group analyses of concept differences were analyzed as one factor in repeated measures analysis of variance for the two studies reported here. In all but the first intrasubject analyses, the constructional approach has been more difficult than the other two for most subjects. The rate at which correct responses were made and the percentage of correct responses were both lower on constructional approach tasks. In addition, although few statistical differences were found between abulia and tau effect, the relative equality of these concepts changed from study to study. These results taken as a whole reflect the need to attend to concept differences.

Throughout these investigations the method selected to deal with concept differences was to simplify the tasks that defined constructional approach. Therefore, a number of attempts were made to redefine and reprogram that concept in the attempt to obtain three concepts of equal difficulty. So far this strategy has not been successful. Thus, it appears that other strategies are worth pursuing.

As stated in the discussion of experiment 1, two alternative strategies are most apparent. The constructional approach could be dropped from further investigations and another concept could be programmed and tested for its similarity to abulia and tau effect. If such a concept was found, this strategy would allow isolating and manipulating one variable, study programs. However, this strategy may not prove to be very satisfactory. It will be recalled that a variety of differences were found between abulia and tau effect across experiments. These differences indicate that even two concepts of apparently similar difficulty do in fact, influence subject performance differentially. Therefore, even if another similar concept could be programmed, subtle differences might not bring these investigations any closer to determining the value of different types of questions. Therefore, a second strategy should be considered.

A second means of handling the concept differences is to look at the interactions between different kinds of concepts and different kinds of tasks. As stated in experiment 1, this implies a change in the direction of future investigations. Rather than trying to determine the study program that contains the necessary and sufficient classes of tasks to promote transfer to a variety of classes of verbal behavior; a more basic investigation of the relations between concepts[3] and verbal skills would be attempted.

For example, it appears that the relations between constructional approach and at least the three kinds of verbal skills programmed here are different from the relations between abulia and these same verbal skills. However, the data presented here do not provide any way to turn these differences into useful strategies for teaching one type of concept as opposed to another type of concept. If intrasubject comparisons are to be made between different concept/task interactions than massive changes in methodology would be required. Concept/task interactions could not be interpreted by the current methods.

One change in methodology would be to make the experimental design more dynamic. By this it is meant that subjects need to interact with each of the different kinds of tasks and concepts more often. More frequent sessions would allow the experimenter to change systematically tasks and/or concepts when subject response patterns were stable. The stability would allow clearer interpretations of the task and concept manipulations than those provided by the current design. In fact, one dependent variable that should be observed is the number of sessions that are required to learn each class of task for a particular concept. In addition, a more dynamic design would allow for a complete factorial analysis. This would eliminate the limitations of a latin square design to interpret interaction effects.

More concepts would also need to be programmed. At this point it is impossible to determine the particular aspects of the concepts that contribute to the differences in subjects' response patterns. However, if other concepts that vary along a number of different characteristics are programmed, perhaps the critical aspects can be determined. This change is also required by the more dynamic experimental design proposed. If many concept manipulations are to be made, then many concepts need to be programmed.

Third, in order to program concepts that vary along a number of different features, it will be necessary to develop some general way of categorizing concepts or concept features. It may be that concepts can be categorized according to certain environmental features that are always present whenever various behaviors are emitted. However, this motion dredges up many issues on defining concept similarity. Can similarity be analyzed for responses and for stimuli? Are response similarity and stimulus similarity or some combination of the two sufficient for categorizing concepts? The answers to these questions are critical for any environmental model of conceptual learning.

Of course, a number of potential answers have been postulated by previous investigators. The most complete suggestions have discussed a continuum of stimulus-response similarity in order to predict transfer (Robinson, 1927;

Gibson, 1940; Osgood, 1949). However, as these were based on explicitely associationist models, they each have problems related to those discussed in Chapter 1 of this manuscript. Specifically, similarity was difficult to describe for responses (Ellis, 1965) and the models did not seem to account for many instances of conceptual behavior (Royer, 1979).

As noted in Chapter 1, the problems of an associationist model of conceptual learning may be minimized by constructing an operant model. Perhaps the problems encountered in defining a concept can also be solved by applying components of a radical behavioral model to a working definition of a concept.

In order to proceed with a radical behavioral definition of a concept a functional definition of a response might be useful. Kantor (1933) emphasized the impossibility of defining a response without reference to a functional stimulus and vice versa. By this it is meant that a response is not simply movement, but the relation between antecedent events and movement. Skinner (1938) added to this analysis by claiming that in order to define a response it was necessary to include the consequence or effect of the movement as well.

This emphasis on the effects of behavior as a critical dimension clarifies the empirical definition of similarity used by operant psychologists. Skinner (1938) emphasized that it was possible to measure the similarity that exists for a class of responses in terms of common effects. However, in order to avoid a completely circular argument, it was also necessary to distinguish between instances of behavior and classes of behavior. Skinner (1938, 1969) reserved the use of the term response to refer to single instances of behavior while the term operant was used to refer to a class of responses. If a class of responses or an operant was to be defined unambiguously, many instances of the relation between behavior and effects would have to occur. If the effects are observed to occur repeatedly, then any movements that are observed to occur concurrently with these effects become possible components of an operant. If movements reliably occur concurrent with particular effects, then the combination of movements and effects is defined as an operant. Since the effect is the common, defining feature, it is not necessary for each possible instance to have occurred in the past, but rather that enough instances have occurred to define a particular effect. Therefore, to return to an earlier statement quoted from Chomsky (1971); if a statement that has never been heard or said before is emitted, it is appropriate to say that it could have been predicted on the basis of its similarity to past events. However, the similarity would have to do with the effect of the statement, not necessarily the exact form or topography of the statement. In the simple case, it could be a general social effect. In the more complex case, it could be a specialized social effect in which only certain words and syntax would have the predicted effect. How often has each reader manipulated the semantic and syntactic structure of a statement to obtain a certain effect with a certain audience? If that audience and the conditions surrounding the audience and speaker could be reproduced, it might be possible to predict the words and syntax used by the speaker.

This definition of an operant allows the observer to make unambiguous predictions about further instances of the operant. This definition can also be applied to the problem of defining a concept. If antecedent, behavior and consequence, are all included in a definition of a concept, it makes little sense to distinguish between concepts and behavior. Behavior is a part of a conceptual operant. If a range of instances of verbal behavior reliably occur concurrently with certain antecedents and certain consequences, then a conceptual operant has been formed.

The single most enduring problem of applying this definition of an operant to conceptual behavior is that the effects that can easily be measured are meaningless (e.g., movement of air molecules for vocal behavior, changes in color of paper for written behavior). The effects that are important are the effects that conceptual behavior has on another individual. But, how can these effects be measured? Part of the answer lies in looking at the behavior of the second individual. For example, in an instructional environment the teacher essentially says either "yes, you are right and here are the reasons" or "no, you are wrong and here are the reasons" or some variation of these (Carlson, 1981). The yes/no response of the teacher is the effect that a student's behavior has had that can be measured. This is fine if an investigator is interested just in the correctness of conceptual behavior. However, there are two reasons that this account alone does little to help define different classes of conceptual behavior: the ubiquitous nature of the general yes/no effect and because more specific instances of yes and no are as difficult to categorize as the responses that produce them.

It is necessary, therefore, in describing different classes of conceptual behavior to emphasize the third term of a functional account, the antecedent. However, having related antecedents to the definition of an operant, the term antecedent takes one slightly different meaning than when discussed under associationist models. In fact, the function of the antecedent in an operant model is different enough to warrant a new term, discriminative stimulus.

A discriminative stimulus is an antecedent that is a component of an operant. That is, if an event reliably occurs before behaviors that have the same effect on the environment, and these behaviors do not occur reliably when the event is not present, then the event is called a discriminative stimulus. Therefore, although the definition of conceptual behavior has turned to antecedent events in order to distinguish between different conceptual operants, the consequences are still critical to the analysis. The stimulus-response association does not stand alone.

The discriminative stimulus of interest in conceptual learning are also the responses of another person. In particular, for study behavior the discriminative stimuli are the questions or tasks set up by the teacher. However, as in all complex cases the questions serve as multiple or conjunctive discriminative stimuli. In other words, the question is a discriminative stimulus for two possible operants; 1) a certain class or sub-class of verbal behavior and 2) a certain class of behaviors related to specific features of the environment. For example, the question, "Describe a cumulative record," is a discriminative stimulus for making a definitional response (one sub-class of verbal behavior) and discussing those features of the environment that are categorized by the term "cumulative record." Thus, in order to assure that an individual will learn to make the appropriate response in the presence of a conjunctive discriminative stimulus, it is probably necessary to teach instances of both operants, and possibly necessary to teach instances of the conjunctive operant.

Since this thesis has described a precise, functional method for defining one of these types of operants, classes of verbal behavior, the remaining task is to integrate this set of descriptions with equally precise descriptions of the features of the environment. A discussion of this issue is beyond the limits of this paper. However, let it be emphasized that the classification of features of the environment should be as functional as the classification of verbal behavior presented here.

Regardless of the massive changes in methodology that have been discussed, this strategy is more appealing than previous strategies because it reflects a progression toward more testable questions and to an experimental analysis of a natural state of conceptual learning. It was argued earlier that although it may be possible to answer questions regarding conceptual learning by developing artificially similar concepts, previous attempts to answer transfer and concept learning questions from this perspective produced reproducible data, but little application to the real world. Therefore, future research will deliberately manipulate the relations between real concepts and study programs.

These changes in methodology would also help to evaluate the problem of interpreting the other effects that were suggested by the current study. Such factors as order of study programs and pretest/no pretest were found to effect some of the dependent variables and not others. Any attempt to describe why such effects occurred would be pure conjecture. The most likely reason is a simple sampling error. However, by developing a more dynamic methodology that includes more data analysis and more concepts, there will be greater opportunity for manipulating order and pretest factors. These manipulations could lead to isolating the reason that order and pretest effects occurred.

Interaction data from the current study are even more difficult to interpret than the simple effects because of the latin square design that was used. Therefore, further manipulations and more data are needed in order to determine how these interactions between order of presentation, pretest and other factors influence the analysis of study programs. Analyzing these interactions could also be continued with the proposed changes in methodology. However, a number of separate studies would be required, each one specifically designed to test specific interactions. If this is done then the problems related to latin square designs will be eliminated.

In summary, it appears that the primary difficulties with interpreting the results of the present study can be alleviated by changing the methods used in future studies. A more dynamic experimental design has been outlined that will make it possible to manipulate the factors that appear to have significant effects on study and test performance.

Thus, a number of concepts need to be selected, programmed and pilot tested to use in the proposed methodology. Then, each of the factors and interactions between factors can be tested with a variety of concepts in several separate, long term studies.

Future Directions - A Summary and Brief Extension. Considerable data have been presented and discussed in this dissertation. These data suggest many tangental studies. Variables that were interesting only as factors to be controlled at the onset of these investigations have generated questions as critical and interesting as the original questions. At one level the data suggest applied studies, at another, more rigorous laboratory studies, and at still another, that more historical and possibly predictive variables be investigated. Certainly, all these directions are worthwhile. In addition, each set of studies can use the information gained from the others. Therefore, this section details some of the questions that can be asked from each of these areas, describes one strategy for integrating all three areas and concludes that the facts and arguments accumulated in this manuscript indicate the importance of all of these investigations.

As indicated in the discussion above, a number of variables appear to affect student performance on tests of different classes of verbal behavior. In order to sort through these variables it is necessary to ask a number of different kinds of questions.

At present it appears that some questions can be best answered in a highly controlled laboratory setting: "Can concepts be classified according to a synthesis of classes of environmental features and classes of verbal behavior?" "Does such a classification system cover the range of conceptual behavior in which adults engage?" "Does such a classification system lead to better control of transfer from one set of learning experiences to other kinds of learning? "Can rules be developed for designing study programs that are efficient for the student to use, yet also facilitate a variety of conceptual behavior?" "Will these rules be dependent on the class of environmental features with which the student interacts?" "Are some classes of environmental features more amenable to some types of tasks than others?" and "Do other factors account for more variation in subject responses than the type of task or the type of environmental features?"

Some historical questions about predictive factors that need to be integrated with those asked above are: "What kinds of student characteristics predict different study and test performance?" "Can evaluation instruments be designed to determine these student characteristics?" "If some students do not require study programs to engage in a variety of conceptual responses, what kinds of study skills do they use and can those study skills be taught to other students through the use of study programs?"

At different points throughout the laboratory and the historical investigations, applied studies need to be conducted to determine the practical validity of the basic findings. After all, what teacher is going to pay students for correct answers or give immediate feedback on all the answers to study questions? In addition, the applied studies may generate potential questions and strategies for further basic investigations. For example, applied studies are needed to answer the following kinds of questions: "If a combination of prose, copy tasks and example identification tasks are sufficient to teach a variety of conceptual behavior in the laboratory, what is the effect of teaching a complete lesson of concepts through this combination of tasks?" "If a number of related concepts are taught in a class, what is the effect of gradually eliminating some types of study tasks?" "How much time does it take to develop a complete lesson of example identification tasks and how much is gained over less time consuming programming strategies?"

Since the questions above tend to look like at least three different research programs, it is important to demonstrate how the three areas of research; laboratory, historical and applied, can be integrated. For instance, a previous section discussed the need to develop study programs and tests for a range of concepts. One way to do this is to begin programming content areas that will be used in real school settings. By directing the programming efforts toward real content areas a number of objectives will be accomplished. First, a range of different kinds of concepts will be developed from which to sample in future laboratory studies. Second, the population of students who will actually use the concepts in both the classroom and in the studies can help develop and test the materials. Third, some initial, on-going tests of the practical validity of these investigations can be conducted. Fourth, data on entering skills should be more readily available on school content areas than on obscure, arbitrary concepts. In addition, other test scores that may predict performance on study programs would be available in schools. Fifth, sets of publishable curriculum materials could be made available to a larger population if certain strategies affect learning better than other strategies. As suggested by Popham (1969), educational practice is more likely to change when the results of research have been made available through curriculum materials. A perfect example of this type of impact is the DISTAR materials developed by Engelmann and his colleagues (cf. Engelmann, 1971; or Becker and Carnine, 1980).

The sum of these objectives should provide enough information to continue the laboratory investigations. These will take the form suggested in the previous section. A dynamic, long-term intrasubject design will be used. This design will allow manipulation of variables when stability of responding has occurred and will also give the experimenter more opportunities to change each variable systematically. Group comparisons will still be made on such factors as order of presentation, pretesting and no pretesting and to determine whether any historical variables are predictive of performance. Thus, in developing the materials from real content, the investigations can flow back and forth between the laboratory and the applied setting.

In addition to these further developments, other classes of tasks should be programmed and tested. Up to this point only three sub-classes of the intraverbal have been manipulated. Perhaps the small differences that have been found between study programs and task types reflect the functional similarity of these categories. If, for example, a program that teaches tact behavior was compared to an intraverbal program of definition tasks, greater differences might be found. Another class of behavior that needs to be investigated is the mand (a verbal response which occurs under conditions of depriation or aversive stimulation and which specifies the reinforcer which will change the conditions for the individual, (Skinner, 1957). The mand appears to be a critical problem-solving and scientific skill. A good problem-solver needs to ask the right questions, at the right time and in the right way in order to bring about a solution. These kinds of questions are directly related to the mand. Therefore, it appears that considerable progress could be made in the area of problem-solving if the mand was investigated.

In summary, it appears from the above discussion that there are limitless studies that can be conducted as extensions of the research reported here. However, are these kinds of studies, which certainly have been proposed before from different perspectives, important and does the functional typology that is the core of all the proposed studies provide a prospective different enough and strong enough to answer these questions. It is hoped that this manuscript has made it clear that both of these questions can be answered in the affirmative.

First, the introduction should have indicated that a model of conceptual behavior that is different from both the associationist and the cognitive models is needed and that current strategies for classifying conceptual behavior are insufficient. It should be clear from this introduction that a radical behaviorial model provides an emphasis on the interactions between the environment and behavior that has not been systematically applied to studying conceptual behavior. Finally, it should be clear that a radical behavioral model of verbal behavior (Skinner, 1957) can provide an alternative model of conceptual behavior that does generate testable research questions. Of course, this model of conceptual behavior is the functional typology described and stressed here (Johnson and Chase, 1981).

In addition, a few simple facts have merged from the initial studies incorporating the functional typology. It is known that subjects can learn to discriminate between different categories of the typology in a reasonable amount of time (Chase, 1980, Johnson and Chase, 1981). It has been shown here that subjects performed differently on types of tasks that are included in a study program than on types of tasks that are not programmed. It has also been shown that three types of tasks tend to promote characteristically different patterns of rate and accuracy and that combining questions in a study program also produces different patterns of responding. Previous studies have indicated that the use of examples and nonexamples facilitated learning (cf., Miller and Weaver, 1976; Markle and Tiemann, 1970; Johnson and O'Reilly, 1964). These results were reproduced here by using the example identification task type in study programs. Not only were example identification study programs more efficient to use, but subjects' performance was as good as after any of the other programs tested.

Finally, the changes in methodology suggested by this discussion chapter should indicate that further information can be gained from the use of the typology. These methodological changes should be stimulating for a number of reasons:

they pose a number of experimental design challenges and a range of studies that may answer previously unanswered questions about conceptual behavior. But now, this summary has returned to its original point: are these studies important and can they generate results that are generalizable and useful to the real world of learning concepts. Only further empirical investigations will supply the answers.

FOOTNOTES

1. All underlined words are defined in the Glossary, Appendix A.

2. This quote is taken from a critique of Skinner's work (Chomsky, 1971).

   However, the discussion of the problems of antecedent stimulus similarity is more appropriate in the context of the associationist model. As usual, Skinner and those who are interested in reinforcement theory are misclassified as associationists. Later, it will be shown how an operant model is different from associationist theory.

3. The reader will recall that an operant definition of concept requires both environmental features and responses (pg. 4 and glossary). Thus, according to this definition each of these relations could be considered a different concept. However, in order to simplify this exposition the term concept is used to categorize features of the environment that are responded to with a single term. Later, the distinction between this over simplified definition of concept and a more complex, thorough definition are discussed in terms of conjunctive conceptual operants.

REFERENCES

American Association for the Advancement of Science. Commission on Science Education Newsletter, 1965, 1, 2-4.

Anderson, J.R. Arguments concerning representation for mental imagery. Psychological Review, 1978, 85, 249-277.

Anderson, J.R. & Bower, G.H. Human Associative Memory. Washington D.C.: V.H. Winston & Sons, 1973.

Anderson, R.C. How to construct achievement tests to assess comprehension. Review of Educational Research, 1972, 42, 145-170.

Anderson, R.C. & Biddle, W.B. On asking students questions about what they are reading. In G.H. Bower (Ed.) The Psychology of Learning and Motivation, (Volume 9). New York: Academic Press, 1975.

Anderson, R.C. & Kulhavy, R.W. Learning concepts from definitions. American Educational Research Journal, 1972, 9, 335-390.

Andre, T. Does answering higher level questions while reading facilitate productive learning. Review of Educational Research, 1979, 49, 280-318.

Andre, T. & Sola, J. Imagery, verbatim and paraphrased questions and retention of meaningful sentences. Journal of Educational Psychology, 1976, 68, 661-669.

Ausubel, D.P. The Psychology of Meaningful Verbal Learning. New York: Grune & Stratton, 1963.

Baer, D.M. Perhaps it would be better not to know everything. Journal of Applied Behavior Analysis, 1977, 10, 167-172.

Becker, W.C. and Carnine D.W. Direct instruction: an effective approach to education intervention with the disadvantaged and low performers. In B.B. Lahey and A.E. Kazden (Eds.) Advances in Clinical Child Psychology, (Volume 3). New York: Prenum, 1980.

Block, J.H. and Burns, R.B. Mastery learning. In L.S. Shulman (Ed.) Review of Research in Education. Itasca, Ill.: F.E. Peacock Inc., 1976.

Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.H. & Kratwohl, D.R. A Taxonomy of Educational Objectives: Handbook I: The Cognitive Domain. New York: MacKay, Inc. 1956.

Bostow, D.E. The Experimental Analysis of College Instructional Practices: Some Mid-Course Corrections. Invited address presented at the Second Annual Mid-Western Association for Behavior Analysis Convention, Chicago, 1976.

Carlson, E. Personal communication. 1981.

Chase, P.N. A Methodology for Evaluating Classification Schemes and Verbal Learning Variables. Unpublished Master's Thesis, University of Massachusetts, 1980.

Chase, P.N., Johnson, K.R. & Keenan, J.B. Survey of commercially available study materials. Unpublished manuscript, 1977.

Chomsky, N. A review of B.F. Skinner's Verbal Behavior. Language, 1959, 35, 26-58.

Chomsky, N. The case against B.F. Skinner. The New York Review, 1971.

Chomsky, N. Aspects of the Theory of Syntax. Cambridge MA: MIT Press, 1965.

Clark, D.C. Teaching concepts in the classroom: A set of teaching prescriptions derived from experimental research. Journal of Educational Psychology, 1971, 62, 253-278.

Deese, J. Psychology of Learning. New York: McGraw-Hill, 1958.

deLuna, S.V. Learning and Transfer in Hierarchically Structured Concept Attainment Tasks. University of Massachusetts, Amherst: Unpublished Master's Thesis, 1972.

Ellis, H.C. The Transfer of Learning. New York: The Macmillan Co., 1965.

Engelmann, S. Conceptual Learning. San Raphael, Ca.: Dimensions Press, 1969.

Engelmann, S. The efectiveness of direct instruction on IQ performance and achievement in reading and arithmetic. In J. Hellmuch, (Ed.), Disadvantaged Child, (Volme 3). New York: Brumer/Mazel, 1971.

Fisher, R. Statistical Methods and Scientific Inference. London: Oliver and Body, Ltd., 1956.

Frase, L. Some unpredicted effects of different questions upon learning from connected idscourse. Journal of Educational Psychology, 1968, 59, 197-200.

Funk & Wagnalls. New Standard College Dictionary of the English Language. C.E. Funk (Ed.) New York: Fund and Wagnall Co., 1963.

Furth, H.G. The influences of language on the development of concept formation in deaf children. Journal of Abnormal and Social Psychology, 1961, 63, 386-389.

Gagne', R.M. The acquisition of knowledge. Psychological Review, 1962, 69, 355-365.

Gagne', R.M. The Analysis of Instructional Objectives for the Design of Instruction. In, R. Glazer (Ed.), Teaching Machines and Programmed Instruction II: Data and Directions. Washington D.C.: National Education Association, 1965, 21-53.

Gagne', R.M. The Conditions of Learning. New York: Holt, Rinehart, and Winston, 1965, 1970.

Gagne', R.M. & Brown, L.T. Some factors in the programming of conceptual learning. Journal of Experimental Psychology, 1961, 62, 313-321.

Gagne', R.M., Mayor, J.R., Garstens, H.L. and Paradise, N.E. Factors in acquiring knowledge of a mathematical task. Psychological Monographs, 1962, 76, (7), 526.

Gagne', R.M. and Paradise, N.E. Abilities and learning sets in knowledge acquisition. Psychological Monographs, 1961, 75, (14), 518.

Gagne', R.M. and staff, University of Maryland Mathematics Project Some factors in learning non-metric geometry. Monographs of Society for Research in Child Development, 1965, 30, 42-49.

Galanter, E. (Ed.) Automatic Teaching: The State of the Art. New York: Wiley & Sons, 1959.

Gallup, H.F. & Warranch, H.R. Some Potential Problems With Commercialization of PSI Materials. In, J.G. Sherman, R.S. Ruskin, & R. Lazar (Eds.), Personalized Instruction in Higher Education Today: Proceedings of the Third and Fourth National Conferences. Washington, D.C.: Center for Personalized Instruction, 1978.

Gerlach, V.S. and Sullivan, H.J. Constructing Statements of Outcomes. Ingelwood, CA.: Southwest Regional Laboratories for Educational Research and Development, 1967.

Gibson, E.J. A systematic application of the concepts of generalization and differentiation to verbal learning. Psychological Review, 1940, 47, 196-229.

Harlow, H. The formation of learning sets. Psychological Review, 1949, 56, 51-65.

Hayes, W.L. Statistics for the Social Sciences. New York: Holt, Rinehart and Winston Inc., 1973.

Health Sciences Computing Facility. Description and Estimation of Missing Data. Los Angeles, CA.: University of California, 1977.

Hersen, M. & Barlow, D. Single Case Experimental Designs: Strategies For Studying Behavior Change. New York: Pergamon Press, 1976.

Hilgard, E.R., Irvine, R.P. & Whipple, J.E. Rote memorization, understanding and transfer: An extension of Katona's card-trick experiments. Journal of Experimental Psychology, 1953, 46, 288-292.

Hirshfield, G.A. Taxonomic Approach to the Evaluation of Secondary School English Classes. University of Illinois: Unpublished Doctoral Dissertation, 1969.

Hively, W. Programming stimuli in matching to sample. Journal of the Experimental Analysis of Behavior, 1962, 5, 279-298.

Hively, W. Domain-referenced achievement testing: Theory and practice. Unpublished manuscript, 1970.

Holland, J.G. and Skinner, B.F. The Analysis of Behavior. New York: McGraw-Hill Inc., 1961.

Holland, J.G. and Kemp, F.D. A measure of programming in teaching machines materials. Journal of Educational Psychology, 1965, 56, 264-269.

Houghton, E. Precision Teaching. Workshop presented at Fernald School, Waltham, MA, August, 1979.

James, W. Principles of Psychology. New York: Holt, 1980.

Johnson, K.R. and Ruskin, R.S. Behavioral Instruction: An Evaluative Review. Washington D.C.: American Psychological Association, Inc., 1977.

Johnson, K.R. and Chase, P.N. Behavior Analysis in Instructional design: a functional typology of verbal tasks. The Behavior Analyst, 1981, 4.

Johnson, K.R. & Chase, P.N. Teaching Concepts, and Operations to High School and College Students. Workshop presented at the Fourth Annual Association for Behavior Analysis Convention. Chicago, Ill., 1978.

Johnson, K.R. & Chase, P.N. Instructional Applications of B.F. Skinner's Analysis of Verbal Behavior. Worktext for Invited Workshop presented at the Fifth Annual Association for Behavior Analysis Convention. Dearborn, Mi., 1979.

Johnson, D.M. & O'Reilly, C.A. Concept attainment in children: Classifying and defining. Journal of Educational Psychology, 1964, 55, 71-74.

Johnson, D.M. & Stratton, R.P. Evaluation of five methods of teaching concepts. Journal of Educational Psychology, 1966, 57, 48-53.

Johnston, J., O'Neill, G.W., Walters, W.M. & Rasheed, J.M. The measurement and analysis of college student study behavior: Tactics for research. In, J.M. Johnston (Ed.) Behavior Research and Technology in Higher Education. Springfield, Ill.: Charles C. Thomas Publishers, 1975.

Johnston, J.M. & Pennypacker, H. A behavioral approach to college teaching. American Psychologist, 1971, 26, 219-244.

Johnston, J.M. & Pennypacker, H.S. Strategies and Tactics of Human Behavioral Research. Hillsdale, N.J.: L. Erlbaum Associates, Inc., 1980.

Kantor, J.R. In defense of stimulus-response psychology. Psychological Review, 1933, 40, 324-336.

Katona, G. Organizing and Memorizing. New York: Columbia University Press, 1940.

Keenan, J.B. & Grant, L. The Question of Relevance and Level of Learning: A Stimulus Control Question. Paper presented at the Fifth Annual Conference on Personalized Instruction, Washington D.C., 1979.

Keenan, J.B. & Grant, L. Transfer and conceptual instructional tasks: A behavioral view. In K.R. Johnson and P.N. Chase (Eds.) Behavior Analysis of Conceptual Learning. Engelwood Cliffs, N.J.: Educational Technology Publications, Inc., In Press.

Keller, F.S., "Goodbye, teacher..." Journal of Applied Behavior Analysis, 1968, 1, 79-89.

Keller, F.S. & Sherman, J.G. The Keller Plan Handbook. Essays on a Personalized System of Instruction. Menlo Park, CA.: W.A. Benjamin, 1974.

Keller, F.S. Definition of Psychology. Englewood Cliffs, N.J.: Prentice-Hall Inc., 1973.

Keller, F.S. & Schoenfeld, W. Principles of Psychology. Englewood Cliffs, N.J.: Prentice-Hall, 1950.

Kintsch, W. The Representation of Meaning in Memory. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.

Kolb, J.R. Effects of relating mathematics to science instruction on the acquisition of quantitative science behaviors. Journal of Research in Science Teaching, 1967-68, 5, 174-182.

Kratochwill, T.R. Single Subject Research, Strategies for Evaluating Change. New York: Academic Press, 1978.

Kulik, J., Kulik C. & Cohen, P. A meta-analysis of outcome Studies of Keller's Personalized System of Instruction, 1979, 34, 307-318.

Lachman, R., Lachman, J. & Butterfield, E. Cognitive Psychology and Information Processing. Hillsdale, N.J.: L. Erlbaum Associates, Inc., 1979.

Lindsley, O.R. 100 Years of Learning Compared on Standard Acceleration Charts. Address presented at 87th Annual American Psychologists Association Convention, New York, 1971.

MacCorquodale, K. On Chomsky's review of Skinner's Verbal Behavior. Journal of the Experimental Analysis of Behavior, 1970, 13, 83-99.

Markle, S.M. Empirical testing of programs. In, P. Lange (Ed.) Programmed Instruction. Chicago: University of Chicago Press, 1967.

Markle, S.M. & Tiemann, P.W. Behavioral analysis of cognitive content. Educational Technology Magazine, 1970, 10, 41-45.

Merrill M.D. & Tennyson, R.D. Teaching Concepts: An Instructional Design Guide. Englewood Cliffs, N.J.: Educational Technology Publications, Inc., 1977.

Michael, J. Statistical inference for individual organism research: mixed blessing or curse? Journal of Applied Behavior Analysis, 1974, 7, 647-653.

Miller, L.K. & Weaver, F.H. A behavioral technology for producing concept formation in university students. Journal of Applied Behavior Analysis, 1976, 9, 289-300.

Morrisett, L.J. & Hovland, C.I. A comparison of three varieties of training in human problem solving. Journal of Experimental Psychology, 1959, 58, 52-55.

Myers, J.L. Fundamentals of Experimental Design. Boston: Allyn and Bacon, Inc., 1979.

Osgood, C.E. The similarity paradox in human learning: a resolution. Psychological Review, 1949, 56, 132-143.

Palmer, D. Notes From a Course on Cognition. Unpublished Manuscript, University of Massachusetts, 1980.

Perkins, M.R. Facilitative Transfer in Prose Learning of Elementary School Children. Unpublished Masters Thesis, 1976.

Peterson, N. An Introduction to Verbal Behavior. Grand Rapids, Michigan: Behavior Associates, Inc., 1978.

Pfeiffer, I.L. Teaching in ability grouped english classes: A study of verbal interactions and cognitive goals. Journal of Experimental Education, 1967, 36.

Polya, G. Mathematics and Plausible Reasoning, Vol. I: Introduction and Analogy in Mathematics, Princeton, J.J.: Princeton University Press, 1954.

Popham, W.J. Curriculum materials. Review of Educational Research, 1969, 39, 319-338.

Postman, L. Transfer, Interference and Forgetting. In, J.W. Kling & L.A. Riggs (Ed.) Experimental Psychology. New York: Holt, Rinehart and Winston, Inc., 1971.

Reed, H.E. Factors influencing the learning and retention of concepts. I. The Influence of set. Journal of Experimental Psychology, 1946, 34, 71-87.

Rickards, J.P. Adjunct postquestions in test: a critical review of methods and processes. Review of Educational Research, 1979, 49, 181-186.

Rickard, J.P. & Denner, P.R. Inserted questions as aids to reading text. Instructional Science, 1978, 7, 313-346.

Robinson, E.S. The "similarity" factor in retroaction. American Journal of Psychology, 1927, 39, 297-312.

Rosch, E. Principles of Categorization. In E. Rosch and B. Lloyd (Eds.) Cognition and Categorization. Hillsdale, N.J.: Lawrence Erlbaum Assoc. Inc. Publishers, 1978.

Royer, J.M. Theories of Learning Transfer. Technical Report No. 79, University of Illinois at Urbana-Champaign: Center for the Study of Reading, 1978.

Royer, J.M. Theories of transfer of learning. Educational Psychologist, 1979, 14, 53-69.

Russel, B. Philosophy New York: W.M. Norton and Company, 1927.

Ryle, G. The Concept of Mind. London: Hutchinson, LTD., 1949.

Semb, G. & Spencer, R. Beyond the level of recall: An analysis of complex educational tasks in college and university instruction. In, L.E. Fraley & E.A. Vargas (Eds.) Behavior Research and technology in Higher Education. Champaign, Ill.: Research Press, 1976.

Sidman, M. Tactics of Scientific Research. New York: Basic Books Inc., Publishers, 1960.

Sidman, M., Cresson, O., and Willson-Morris, M. Acquisition of matching to sample via mediated transfer. Journal of the Experimental Analysis of Behavior, 1974, 22, 261-274.

Sidman, M. and Stoddard, L.T. The effectiveness of fading in programming a simultaneous form discrimination for retarded children. Journal of the Experimental Analysis of Behavior, 1967, 10, 3-15.

Skinner, B.F. Are theories of learning necessary? Psychological Review, 1950, 57, 193-158.

Skinner, B.F. The Behavior of Organisms. Englewood Cliffs, N.J.: Prentice-Hall Inc., 1938.

Skinner, B.F. Verbal Behavior. Englewood-Cliffs, New Jersey: Prentice-Hall, Inc., 1957.

Skinner, B.F. Teaching machines. Science, 1958, 128, 969-977.

Skinner, B.F. Why we need teaching machines. Harvard Educational Review, 1961, 31, 377-398.

Skinner, B.F. Contingencies of Reinforcement. Englewood Cliffs, N.J.: Prentice-Hall Inc., 1969.

Stoddard, L.T. and Sidman, J. The effects of errors on children's performance on a circle-ellipse discrimination. Journal of the Experimental Analysis of Behavior, 1967, 10, 261-270.

Sullivan, H.J. Objectives, evaluation, and improved learner achievement. In, J. Popham, E.W. Eisner, H.M. Sullivan & L.L. Tyler (Eds.) Instructional Objectives. AERA Monograph Series on Curriculum Evaluation, #3, Chicago: Rand McNally, 1969.

Trehub, A. Neuroual models of cognitive processes: networks for learning, perception and imagination. Journal of Theoretical Biology, 1977, 65, 141-169.

Watts, G., and Anderson, R.C. Effects of three types of inserted questions on learning from prose. Journal of Education and Psychology, 1971, 62, 387-394.

White, R.T. Research into learning hierarchies. Review of Educational Research, 1973, 43, 361-375.

White, O.R. & Haring, N.G. Exceptional Teaching. Columbus, Ohio: Charles E. Merrill Co., 1976.

Williams, R.G. A behavioral typology of educational objectives for the cognitive domain. Educational Technology Magazine, 1977, 17, (6).

Winokur, S. A Primer of Verbal Behavior: An operant view. Englewood Cliffs, N.J.: Prentice-Hall, Inc. 1976.

Wolff, J.G. Effects of verbalization and pretraining on concept attainment by children in two mediation categories. Journal of Educational Psychology, 1967, 58, 19-26.

APPENDIX A

Glossary

1. Complex verbal behavior (complex conceptual behavior) - verbal or conceptual behavior that includes mastery of a variety of different classes of verbal behavior. Therefore, instances of verbal behavior that incorporate two or more classes of verbal behavior. (Johnson and Chase, 1981).

2. Conceptual behavior - verbal behavior that is flexible or that occurs in the presence of novel features of the environment and/or that combines instances of verbal behavior in a novel way. (Johnson and Chase, 1981).

3. Concept - those features of the environment that are responded to or reacted to similarly. Similarity can range from using the same term to categorize features of the environment to complicated responses such as a definition. A concept is always defined in terms of behavior and therefore, there must be repeated instances of any features of the environment that are part of a concept. (Keller, and Schoenfeld, 1950), (Johnson and Chase, 1981).

4. Critical features - components of a set of environmental events that must be present in order to classify the set in a specific way. The components that make up the stimulus aspects of a concept. (Markle and Tiemann, 1970).

5. Elementary verbal behavior - verbal behavior that is rigid or fixed. Each instance of behavior is identical and occurs in the presence of the same stimuli. (Johnson and Chase, 1981)

6. Example - an instance of the environmental events that comprise a concept. Usually an example is an illustration or a concrete description of an instance of a concept.

7. Extension tasks - items, tasks or questions on a probe or test that are novel instances of classes of tasks that have been explicitly trained before. (Johnson and Chase, 1981).

8. Hierarchy - an order of persons, objects or events classified according to rank or arranged in successive classes each of which is subject to or dependent on the one above it. (Funk and Wagnall, 1963).

9. Nonexample - an instance of environmental events that does not contain <u>all</u> of the features necessary to categorize a concept. Usually an illustration of another concept that is similar to the concept being taught, but different in some critical way.

10. Tasks, items, questions - the verbal behavior that a teacher emits which explicitly requests a response from a student.

11. Taxonomy - a classification system that is arranged in categories beginning with the broadest and most inclusive and ending with the narrowest. (Funk and Wagnall, 1963).

12. Teacher - all forms of instruction including parents, administrators, trainers, textbook or any other learning materials that are accessible to the student.

13. Transfer tasks - classes or types of items, tasks or questions on a probe or test that have not been explicitely trained before.

14. Typology - a classification system that defines each type or category or class without specification of order.

15. Verbal behavior - any response which is effective only through the mediation of another living organism. (Skinner, 1957).

16. Verbal learning - the acquisition of responses that require the mediation of other organisms. (Skinner, 1957).

APPENDIX B

A Functional Typology of Verbal Tasks, a Classification System

The Classification of Verbal Tasks

Johnson and Chase (1979, 1980, 1981) have designed a typology of verbal behavior that conforms to the rules for defining cognitive behavior derived in the first section. This functional typology was originally based on Skinner's (1957) analysis of verbal behavior. Skinner's functional classification system is a viable theoretical model for integrating language in general. However, strict adherence to this model may cause some practical problems. All of Skinner's terms (i.e. mand, intraverbal, tact, transcriptives) are neologisms. Although Skinner's justification for adopting these terms was well founded (i.e. to eliminate confusion arising from the mentalistic use of traditional labels for cognitive behavior) these terms are new to most practitioners and are not readily assimilated into everyday use. In addition, these terms are not always descriptive of what they represent. In fact, often their descriptive quality begets errors. For example, the mand was derived from such words as demand, command and mandatory. However, demands and commands are not always considered mands. They are often tacts or intraverbals. Such limitations, when coupled with the fact that Skinner's analysis has never been experimentally validated, warrant the changes that have been made here. Thus, the following definitions of cognitive or verbal behavior are derived from

Johnson and Chase (1978, 1979) and Skinner (1957), but the labels have been changed to avoid the confusion inherent in Skinner's terminology. A complete version of the typology is presented in Table 1. As each class of behavior is introduced the descriptive label is followed by the class of verbal behavior from which it is derived.

Two additional requirements have been added. The reader will notice in Table 1 that the categories are divided between elementary and conceptual tasks. Elementary tasks are those that require memorized performance. They are rigid, inflexible and do not involve novel situations. Thus, whenever a student repeats a task exactly as it was performed before, he/she is engaging in elementary behavior. Conceptual tasks are those that are extended to new situations or instances. Conceptual behavior are flexible. Thus, whenever a student completes a new task or answers a novel question, he/she is performing conceptually.

The first class of behavior defined for this typology is the copy task (transcriptive behavior). The student reads a passage that defines, describes or exemplifies a phenomenon. Then, the teacher presents a task that either explicitly asks the student to copy, reproduce or reconstruct the passage (or some subset of the passage) or implicitly increases the likelihood of copying by including sequences of words taken directly from the passage. The student's response duplicates the passage or portions of the passage.

For example:

The student reads the following passage:

Abulia is a term used to describe low rates of behavior that are caused by an abrupt change from frequent to infrequent reinforcement. Sometimes the number of times that we must perform a behavior before reinforcement occurs is too large to keep us behaving. The result is a decrease in the frequency that we engage in the behavior. Freud called the resulting low frequency of behavior abulia.

Then, the teacher presents the following task:

Complete the following sentences by finding the sentences in the passage and copying the appropriate words in the blanks.

1. _______ is a term used __ __________ _____ _______ _____ ________ ______ _____ ______ _____ ___ __________ _____ from __________ to _____________ _________________.

2. Freud called _____ _______ ____ ______ ___ _________ ________.

The student responds by filling in all the blanks from the text above.

Notice that the student and the teacher have used words and sequences of words that are identical to those used in the prose passage.

One nonexample of a copy task is presented below:

The student reads the same passage defining abulia.

The teacher asks the student to fill in the blanks in the following task:

1. Abulia occurs when a ________ or dense ____________ __ reinforcement ______ _______ to a more ____________ or ________ schedule _____ ____________.

Notice the words presented by the teacher in this fill in task are completely different from those used in the prose passage. Therefore, this task makes it very unlikely that the student can copy from the passage.

Another example of a copy task is exemplified below:

The student reads the same passage presented previously. The teacher asks: "What term would you use to describe low rates of behavior that are caused by an abrupt change from frequent to infrequent reinforcement?" The student answers: "Abulia."

This is a copy task because the teacher's presentation includes words or sequences of words taken directly from the passage. The student needs only to look through the passage until he comes across these words and copy the related term from the sentence.

A <u>non</u>example that is similar in structure, but different in function to the copy task just presented is:

After the student reads the passage defining abulia, the teacher asks: "What term would you use to describe the following situation?"

> Julia found that every time she kissed Allen on the neck he would respond by taking her into his arms and kiss her passionately. As a result she would often kiss Allen on the neck. One dark Thursday, Julia kissed Allen on the neck 6 times and he continued to read the paper until the sixth kiss. Then he turned and gave her a big kiss. The next day Julia kissed him 10 times before he responded. After this, Julia rarely Kissed Allen on the neck.

The student responds: "Abulia."

In this case, the student could not merely look back at the passage and find the answer. He has to apply the words presented in the passage to determine if the situation exemplifies abulia or some other concept.

A second nonexample that is structurally similar to the copy task presented above is:

> After reading the passage defining abulia, the student is asked: "What psychological concept refers to decrements in behavior that result from rapid decreases in the schedule or amount of reinforcement that occurs for that response?" The student responds: "Abulia."

Again, the student can not simply reread the prose passage and copy the term abulia. The student must be affected by the defining features of the term and affected similarly by the synonymous sequence of words used in the task.

To summarize, a copy task is characterized by either the necessity for or the possibility that a student will copy sentences or phrases directly from instructional materials. If a task sets the occasion for copying from instructional materials and a copy response could be a correct response, then the task is a copy task.

The second class of behavior that is studied in this project is the _define_ task (intraverbal behavior). The students read or hear a prose passage that defines, describes and/or exemplifies a phenomenon. Then, the teacher presents one or more terms, definitions, rules or partial definitions concerning the passage. If a term is presented, the teacher requests a novel (not previously described) verbal response. Students define or describe the term in their own words. If a definition, rule or partial definition is presented, the teacher uses words or sequences of words that are _not_ used in the passage. The student identifies the definition with a term.

A task category that is easy to confuse with define tasks is the copy task. Thus, a nonexample of a define task is:

> After reading that glabrous skin is defined as skin devoid of hair, the student is presented with the following task: "Define glabrous skin." The student responds: "Skin devoid of hair."

Notice that the student's response is _not_ novel. He has repeated the exact sequence of words provided by the instruction. Notice, also, that the task did _not_ ask for a novel response. This is an example of a copy task, _not_ a define task.

This particular illustration could be altered to make it a define task by simply inserting the phrase "in your own words" at the end of the request. If the student responded: "Glabrous skin is skin that has no hair," then a define task has been completed.

Another example of a define task is:

After reading the definition of glabrous skin, the student is asked the following question: "What term is used for skin that is hairless?" The student responds: "Glabrous."

This is an example of a define task because it requires the synonymous effect of "devoid of hair" and "hairless." In other words, the task uses words that are different from those used in the definition.

Another example of a define task is:

After reading two passages, one defining glabrous skin and one that defines pubescent skin as that covered with short or downy hair, the student is asked: "Compare and contrast the terms pubescent and glabrous skin. Do not repeat the definitions given in the text." The student answers: "Both terms are used to describe different kinds of skin. Glabrous refers to skin or sections of skin that are hairless. Pubescent refers to skin that has small, soft, often hard to see hair."

Again, this is an example of a define task because it explicitly asks for sequences of words different from those provided by the teacher and the student has responded as directed.

Another task category that is easily mistaken for define tasks is illustrated below:

After hearing a prose passage that defines glabrous skin, the student is asked to say which of the following is an example of glabrous skin:

a. the palms of the hand

b. the forearm

c. the soles of the feet

d. lips

e. dorsal side of toes

The student identifies a., c., and d., as glabrous skin. Notice that whereas the words used in this task are different from those used in the definition, they are not terms, definitions, rules or partial definitions of the phenomenon. Rather they are concrete instances of the phenomenon. Thus, this is a nonexample of a define task.

An example of a define task that is similar to the non-example presented above is:

After hearing the definition of glabrous skin, the student is presented the following question:

Say which of the following defines glabrous skin:

a. skin that has soft, downy hair

b. skin that is completely hairless

c. skin that has a hair, no hair pattern

d. skin that is covered with course hair

The student identifies b. as the definition.

Again, the words used in the task are different from those presented in the passage (pg.) and the choices are all rules or general descriptions of skin type. They are not concrete instances of skin. Therefore, this is an example of a define task.

In sum, the define task is defined as the presentation of words, terms or definitions to which the student must respond in his own words. If the student's response is made in general terms or the statement of a general rule, and this response is novel, then the task is a define task.

The third class of verbal behavior categorized by this typology is the exemplify task (intraverbal behavior). The student reads or hears a prose passage that defines, describes and/or exemplifies a phenomenon. Then, the teacher asks the student to give an original example of the phenomenon or some subset of the phenomenon. The student's response is a concrete narration of a novel (not previously described) instance of the phenomenon. The student's narration includes properties of the environment that are irrelevant to the definition of the phenomenon.

This task category is relatively easy to identify. In all cases, it requires that the teacher explicitly request some original description of a concrete instance of a general rule, prediction or definition. For example:

The student reads or listens to a passage that discusses hygrometers. A hygrometer is defined as a device for measuring moisture in the air. An example is given that describes a wet and dry bulb hygrometer. The end of one thermometer is wrapped in cloth, the end of the cloth is extended down into a bottle of water. This hygrometer measures the relative humidity by measuring evaporation. The more water that evaporates, the less moisture that is already in the air. The teacher asks the student to give an original example of a hygrometer. The student writes: "At home we have a carved, wooden Swiss Chalet that houses a boy with an umbrella and a girl in a bathing suit. Each figure is standing on opposite ends of a swivel post that is attached to a tautly stretched human hair. Whenever there is a lot of moisture in the air, the boy swings out of the Chalet. Whenever there is little moisture in the air, the girl swings out." As the prose passage does not contain such a description of a hygrometer, this is an example of an exemplify task. The task explicitly asks for an original example. The student's response is a description of one instance of a hygrometer. It is certainly an original description in relation to the instruction.

One <u>non</u>example of an exemplify task is presented below: The student reads a chapter that describes hygrometers and barometers as instruments for measuring weather change. The teacher asks the student to give an original description of the similarities and differences between hygrometers and barometers. The students answers: "Both barometers and hygrometers are devices for observing changes in climate. The barometer measures changes in air pressure and the hygrometer measures changes in the amount of moisture in the air."

In this case, the teacher has not explicitly asked for an original example of an original description of one instance of the terms hygrometer and barometers. The student has answered the request in terms of general descriptions. She has stated three rules that relate and distinguish between hygrometers and barometers. Thus, this is a define task.

Of course the teacher could change the task above into an exemplify task by substituting the word "example" for "description" or by adding the words "by juxtaposing instances of each" at the end of the request. These changes require that the student answer the request with a concrete narration of instances of both a hygrometer and a barometer.

At this point it seems necessary to specify one critical difference between define tasks and exemplify tasks. Often instructional tasks do not explicitly state the kind of behavior in which the student needs to engage. In these cases, it is necessary to observe the student's response in order to determine the task category. For instance, in the illustration of hygrometer and barometer above, the student could have answered the original question by juxtaposing novel examples of both hygrometers and barometers. He could have related these two instances by making two examples similar except for the specification that one measured air pressure and the other, water content in the air. The student would have correctly answered the teacher's task, however, the answer may not have been what the teacher expected. This original question was ambiguous. It seems clear that in such cases the teacher should rewrite the task, making explicit what is expected from the student.

One last nonexample of exemplify tasks should be sufficient. In this illustration, the student reads the passage defining and exemplifying hygrometers. Then, the teacher asks for an example of a hygrometer. The student repeats the example given by the teacher by stating:

> Take two thermometers, wrap one end of one of the thermometers in cloth and let one end of the cloth extend into a jar of water. This will measure the amount of evaporation and thus, the amount of water in the air.

The problem with this task is that the teacher did not ask for an original example. Therefore, the student has answered the task correctly by repeating the example given in the passage. Again, if the teacher wants an exemplify response, the task must be explicit. In this case, the student's behavior constituted copy behavior.

In conclusion, the exemplify task is characterized by a request to give an original example of a concept. If such a request is answered by a concrete description of one or more instances of a general rule, term or definition, then the task is an exemplify task.

The fourth class of behavior of the functional typology is the example identification task (intraverbal behavior). The student reads or hears a prose passage that defines, describes and/or exemplifies a phenomenon. Then, the teacher presents one or more concrete narrations of novel instances of the phenomenon and/or novel instances of other phenomenon and asks the student to identify those instances that illustrate the phenomenon. The student identifies those descriptions that conform to the general rule of definition of the phenomenon given in the passage.

For example:

A student reads the following passage:

> The constructional approach is a relatively new way by which we can change the problem behavior of an individual. Currently, most methods for dealing with problem behavior focus on eliminating or alleviating the distressing behavior. An alternative, the constructional approach, focuses on teaching new behaviors that are followed by desirable outcomes. This is accomplished by determining the desirable outcomes that maintain the problem or distressing behavior. Then, the constructional therapist helps construct alternative behaviors that are maintained by the same critical or desirable outcomes, but that are not distressing to the client.

Then, the teacher presents the following illustrations with the request: "Circle the letter corresponding to each of the following that is an example of the constructional approach."

a. Bob enjoyed jogging because of the exhilarating feeling and because it kept him in shape. However, when he jogged in the winter, he got a sore throat, cold feet, and the chills. Since he didn't like these effects, but liked staying in shape, Bob was faced with a dilemma. After hearing Bob complain about this predicament for the hundreth time, his friend Mara suggested that he join a health club and work out on the Nautilus machine. Bob decided to give this a try. Bob no longer jogs every day; he works out on the Nautilus instead. This keeps him trim and he no longer complains about his winter colds.

b. Gerry was a music major participating in a research study conducted by one of the music theory graduate students. Every five seconds, a note was played and Gerry was required to indicate the name of the note. The first note was "B", which Gerry correctly identified. He incorrectly identified the next note as an "F"; it was really another "B". The third note was an "F", and he got that right. The next two notes were "C's"; he identified each of them correctly. Finally, he was presented another "B", which he identified correctly.

The student identified a. as an example of the constructional approach.

Notice that each of these descriptions is concrete and can be considered an instance of some phenomenon. Illustration a. does conform to the definition of constructional approach given in the passage.

A nonexample of an example identification task is: After reading the same passage defining constructional approach, the student is presented the above illustrations as examples and nonexamples of the constructional approach. The student answers: "Example a. does illustrate constructional approach because a new behavior is learned that is reinforced by the same events, exhilaration and trimness, as jogging, but does not have the same bad effects as jogging in the winter. Jogging has also decreased."

In this case, the student does have to identify which is an example of the phenomenon, but he also has to justify his answer in terms of a general rule. Thus, this task is a define task as well as an example identification task.

Another nonexample is:
The teacher asks the student to define constructional approach in his own words. The student responds: "The constructional approach is a model of therapy that attempts to construct new behavior rather than get rid of old, unwanted behavior. The therapist tries to determine what are the positive and the negative consequences of a problem behavior. Then he teaches the

client another behavior that gets the positive consequences, but does not get the negative consequences."
Of course, this is an example of a straight define task. The student hasn't had to identify any concrete narrations. He has provided a definition in his own words.

We could also ask the student to give an original example of the constructional approach. This would not be an example of an example identification task, it would be an example of an exemplify task. Here, the student is providing the novel narration; in the example identification task the teacher is providing the novel narration.

One last point about example identification tasks. The narrations that the teacher presents need to be different from those that are presented in instruction. If an illustration is presented in a passage and then presented as a task, the student need only to look back at the passage, find the narration and copy the term that is related to it. This would be an example of a copy task. Therefore, a task is identified as an example identification task only when the student is presented with a novel description of an instance of a concept and must identify it as such.

The final class of behavior that is of interest to this project, the combination task, is not a single class of behavior at all. Rather, it is a class of tasks that is composed of various combinations of the previously defined classes. The student reads or hears a prose passage that defines, describes and/or exemplifies a phenomenon. Then the teacher asks the student to engage in two or more kinds of tasks with respect to the phenomenon. The student's response is any combination of copying, defining, exemplifying and/or example identifying that is called for with respect to the phenomenon.

For example, the student reads a passage that describes the procedure of effective imitation training. She is then presented with the following task:

Ada was a little-league baseball coach. To help the children learn to field groundballs, she had them hit balls to her while she demonstrated the essential elements of fielding. These elements included getting her body in front of the ball, kneeling on one knee, keeping her eye on the ball, and so on. Then she hit some balls to the children and had them try to field the balls. When they fielded the ball correctly she praised them.

If the above is an example of effective imitation training procedures, key the components of the illustration to the components of your definition of imitation procedures.

The student answers by writing that the illustration does conform to the definition of effective imitation procedures. Ada has modeled the appropriate response, has had the players attempt the response and has given the players feedback on whether or not they are correct. This task clearly exemplifies a combination task. First, the student has to say whether or not the narration describes an instance of effective imitation procedures. This is the example identification component of the task. Second, the student has to generate a definition of effective imitation procedures. This, of course, is the define component. The task is a define/example identification combination task.

This example could be changed to become an example of a single task category very easily. We could make it a define task by deleting the description of "Ada" and instead stating the general rule for effective imitation procedures. We could make it an example identification task by simply asking the student to say whether or not the "Ada" illustration exemplified effective imitation procedures.

Another example of a combination task is illustrated below: The student studies the relations among the demographics of the people in a voting precinct, the kinds of political candidates that have been elected in different communities and the probabilities of new candidates being elected given certain perspectives on issues. Then, the student is given the following task:

> Below are biographical sketches of the residents in three precincts. Accompanying each sketch is a detailed description of successful and unsuccessful candidates in prior elections. Describe the candidate who you could predict would be the most likely to win an upcoming election. Say why you have chosen such a candidate.

Notice that if the sketches and descriptions had been provided then the student would have to engage in three classes of behavior in order to answer this question completely.

The student has to identify political trends from concrete narrations, an example identification task; the student has to describe the characteristics of a candidate most likely to win, an exemplify task; and the student has to justify his descriptions in terms of the general rules of political relations discussed in the text, a define task.

Again, a combination task is any task that asks the student to engage in two or more classesof behavior.

APPENDIX C

Various Examples of Tasks Used in Study Programs and Tests

Copy Task:

Complete the following sentences by finding the sentence in the text and copying the appropriate words in the blanks.

1. To test for the tau effect the _____ _____ _____ _____ _____ _____ _____ _____ or _____ times.

2. If the person _____ _____ _____ or _____ _____ _____ _____ _____ _____ _____ _____ is _____, and _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ is _____, then _____ _____ _____ _____ _____ _____.

3. In fact, research has shown that _____ _____ _____ _____ _____ _____ _____ _____ _____ _____, the _____ _____ people _____ _____ the _____ _____ _____.

Example Identification Tasks:

1. Scott collects rare coins. On his birthday a friend, Artie, gave him a 1909 VDB penny. When his father gave him another one he told Artie that he got two 1980 VDB's for his birthday. Four months later, on Christmas, his Aunt Theresa also gave him a 1909 VDB. His mother asked him if that was the same coin his father had given him for his birthday. He said no, that his father had given him a 1931-S.

44. For one full week, Laura and Jerry talked a lot about the kinds of trees they saw on their way to work. Jerry taught Laura the difference between maple and oak trees. One Wednesday morning, Laura could identify each maple tree they saw. By Thursday morning Laura could identify each maple tree but was still having trouble identifying each oak tree. On Friday morning, Laura could successfully identify each maple tree but still made some mistakes identifying each oak tree they encountered.

Definition Tasks:

1. Two identical objects are presented to a person in rapid succession. How will he respond when asked to compare what he saw during the two presentations?

5. A person is asked to observe a certain event. Much later he is asked to observe the same event. How will he respond when asked to compare the two events that he observed?

12. Describe the tau effect in your own words. Be sure to explain two ways that time determines whether an observer will say that two identical objects or events are the same or different. Be complete.

Terminal D. Define the "tau effect" in your own words.

Exemplify Tasks:

Ex. #1 Given an original example in which 2 identical objects or events are presented to a person with a short period of time between presentations. Have person say that the 2 objects or events are identical.

Ex. #3 Give an original example of the tau effect. Make sure there are at least 3 presentations of similar or identical objects or events; that one time period between presentations is short and one long; have a person identify the objects or events as the same when the time period is short, and finally have a person say that the objects or events are different when the time period is long.

Terminal Ex. Given an original example of the "tau effect."

Combination Task:

10. Kim was reading the newspaper the other day when she noticed a little blurb about Carter's energy policy. The next day, she saw another article on Carter's energy policy, and told her friend Carl that there had been two of the same articles in two days about Carter's energy policy. A month later, in a Statement-of-the-Union address, Carter stated his energy policy but added some features to it. Kim told Carl how she was glad to hear that Carter has expanded his energy policy.

Say whether the above is an example of the tau effect. If it is, justify your answer. If it is not, rewrite the passage to make it illustrate the concept. Then justify the changes you made.

Combination Task:

APPENDIX D

Study Behavior Questionnaire for Experiment 1

STUDY BEHAVIOR QUESTIONNAIRE

S# ______________________

______________________ (name)

______________________ (date)

During study:

1. Please indicate the total time it takes you to read the material for the first time, to the nearest minute: ______________________.

2. Please indicate the amount of time you spend rereading any of the material, to the nearest minute: ______________________.

Indicate the amount of time you spend making any of the following written study aids, each to the nearest minute:

3. outlining: ______________________

4. underlining, brackets, or marking any other parts of the material: ______________________

5. summarizing: ______________________

6. terms, names, definitions, important points, etc.: ______________________

7. examples, applications, uses, etc.: ______________________

8. fill-in questions: ______________________

9. multiple-choice questions: ______________________

10. short or long essay questions: ______________________

11. other: ______________________

12. other: ______________________

13. other: ______________________

14. other: ______________________

15. other: ______________________

16. Indicate the amount of time to the nearest minute you spend orally quizzing yourself with any written study aids you made: ______________________

17. Indicate the amount of time to the nearest minute you spend silently quizzing yourself with any written study aids you made: ____________

18. Indicate the amount of time to the nearest minute you spend writing answers to any written study aids you made: ____________

19. Indicate the amount of time to the nearest minute you spend orally quizzing yourself without your written study aids: ____________

20. Indicate the amount of time to the nearest minute you spend silently quizzing yourself without your written study aids:

21. Indicate the amount of time to the nearest minute you spend writing answers to any nonwritten questions you ask yourself: ____________

2. Indicate the amount of time you spend filling out this questionnaire: ____________

After Study

1. Please indicate your level of disturbance during the study episode by rating any visual, auditory, or other distractions:

1	2	3	4	5
low		medium		high

2. Please indicate the time of the entire study episode to the nearest minute: ____________

3. Please indicate your accuracy in filling out this questionnaire:

1	2	3	4	5
low		medium		high

APPENDIX E

The Pretest for Experiment 1

PRETEST

Read each question carefully. Answer each question as completely as you can. If you don't know the answer to a question, write DK (for don't know).

1. Define "constructional approach" in your own words.

2. Give an original example of "abulia."

3. Say which of the following are examples of "tau effect:"

   a. Todd saw a picture of Cheryl Tiegs modeling a bathing suit in Sports Illustrated, one day in 1973. When he saw her on Cover Girl makeup ads two years later, he told his girlfriend that he had predicted that she would make it big some day. Later that day he saw her ona talk show, and commented on how lucky he was to be able to see her twice in one day.

   b. Cindy likes bikini bathing suits on men. While at the beach one afternoon last summer, she saw a guy with a blue bikini. Two weeks later, she saw another guy wearing a blue bikini, and pointed out to her friend Liz that she had seen the same suit two weeks ago and thought that it would look good on her friend Joe. When she went to the beach again the next day, Liz pointed out the same blue bikini on someone else, but Cindy said the suit she liked is light green.

   c. Mary and Rod went to see "For Pete's Sake," starting Robert Redford. The next day, they went to see "Butch Cassidy and the Sundance Kid." Rod pointed out that Robert Redford was in both movies. Three months later, they went to see "All the President's Men," which also starred Robert Redford. Mary asked, "Isn't that Robert Redford?" Rod replied, "Nah, that's Paul Newman."

   d. Teddy likes to buy crackerjacks. He gets the prizes and then gives them away to his friends. Once he got two blue secret decoder rings in the same box. He went running to his mother and told her that he got the same prize twice. He gave one of the rings to his friend Joe. Three weeks later he was visiting Joe and asked him where he got the blue secret decoder ring. Joe told Ted that he had given it to him. Joe said, "I thought I had given you a red one."

e. For one homework problem, Lisa had to factor the equation $4X^2 + 6X + 4$. In class the next day, the students had a quiz on which that problem was included. She solved the problem, and thanked the teacher for putting the homework problem on the quiz. Two weeks later, on an exam the same problem appeared again. After the exam was over, the teacher laughingly asked Lisa whether any of the problems looked familiar. Lisa said that she didn't think so.

f. Jacques saw the painting "Crows Over a Wheatfield" by Van Gogh, in his art history class, and then later that night, while thumbing through Newsweek he told his friend Sharon, an art major, that he really liked it. Later that year, Sharon and he were walking through the Lincoln Center For the Arts when Sharon pointed out "Crows Over a Wheatfield." She asked Jacques to name the painting. He said, "It's called Country Spring."

4. Roberta is an excellent tennis player. She feels great after returning a difficult serve or hitting a shot right down the line. She is usually able to make a number of such excellent shots in a game. Recently, she has started to spend more time hiking, and has found it difficult to arrange time for tennis. She has gone from playing four times a week to once a week.

Say whether the above is an example of abulia, constructional approach, or tau effect. If it is one of these three concepts, justify your answer. If it is not one of these three concepts, rewrite the passage to make it illustrate the concept to which it is closest. Then justify the changes that you made.

5. Give an original example of the "tau effect."

6. Explain the difference between effective reaction potential and momentary effective reaction potential. Be complete.

7. Define abulia in your own words.

8. Say which of the following are examples of the "constructional approach":

   a. Bob was a top student in high school, although he rarely studied. Things are a little different now that he has come to college. Every night after supper, the guys talk him into playing cards. Often the card games lead to philosophical discussions, which he admittedly loves. As a result, he spends every night playing cards for about three hours, and then expounding upon philosophical issues for another three hours. This leaves little time for homework. After getting a 1.3 cum last semester, Bob decides to see his advisor for suggestions. Together they decide that since he likes philosophizing so much, it might be best for him to arrange his schedule so that he studies right after supper, and discusses philosophy only after he has completed at least 90% of his homework. Since this arrangement, he has handed in nearly all of his homework on time, and has gotten above 90% on all of his exams and quizzes.

   b. Arthur disrupted his third grade class by swearing a great deal. His teacher decided have a talk with him and the principal. They concluded that the shocked face of the teacher and the giggles of the class were encouraging Arthur to swear. The teacher talked to the class and asked them not to giggle anymore. She informed them that if they giggled when Arthur swore, they would not be able to have recess. She also tried not to look shocked anymore. This program effectively reduced Arthur's swearing; he has not sworn in class for overeight weeks.

   c. Greg was mentally retarded and participated in a workshop to learn to assemble transistor radio parts. He spent much of his time distracting the other members of the workshop. They often enjoyed the distraction, and kidded around with him. This caused the supervisor to believe that the kidding around supported Greg's distractive behavior. The supervisor decided that Greg should only be allowed to kid around with the other participinats if he was on task for 15 straight minutes. This proved to be much better for all concerned, as Greg quickly learned to work diligently for 15 minutes and kid around for 5.

d. Mrs. Johnson thought that Shawn talked too much in class. She decided to put an end to this verbosity once and for all. So Mrs. Johnson had the principal call Shawn down to the office to run an errand. Then, she told the class that they were to stop attending to Shawn when she talked for more than two minutes straight, and when she talked more than three times within each class period. She asked them to make sure that they continued to acknowledge Shawn's comments up to these points. Before the end of 4 weeks, Shawn was contributing to the class with maximum effective-ness.

e. Gail and Darcy liked to go for a drink after work on Fridays. They find this to be a time to unwind and catch up on gossip. They frequently went to a local singles bar, but were constantly bothered by men looking for a pick up. It seemed that 9 out of 10 times that they went to a singles bar, they would get approached by obnoxious men. A friend of theirs suggested that they go to a gay bar. Gail and Darcy found that at a gay bar they could still enjoy each others company without being harassed by other customers.

f. Frank spent much of his time isolated in a corner. This bothered his parents, who would like to see him interact more with other people. They decided to talk to a psychologist. He determined that Frank was afraid of other people calling him names and picking on him. He instructed Frank's parents to give Frank money toward a bicycle every time he interacted with others. Because the psychologist did not specify the exact nature of the interactions, Frank got money for calling people names and verbally assaulting them. For the last three months Frank has not sat in a corner once.

9. Give an original example that illustrates how stimulus generalization affects effective habit strength.

10. Kim was reading the newspaper the other day and noticed a little blurb about Carter's energy policy. The next day, she saw another article on Carter's energy policy and told her friend Carl that there had been two separate articles in two days about Carter's energy policy. A month later, in a STATEMENT-OF-THE-UNION address, Carter expounded further on his energy policy. Kim told Carl how she was glad to hear that Carter has expanded his energy policy.

    Say whether the above is an example of abulia, constructional approach, or the tau effect. It it is one of these three concepts, justify your answer. If it is not one of these three concepts, rewrite the passage to make it illustrate the concept to which it is closest. Then justify the changes that you made.

11. Georgia, a three year old attending a local preschool, had a history of whining when things didn't go her way. For example, one day when she was building a house out of blocks, she accidently knocked it over. She immediately started whining, and her teacher came right over. This sort of behavior had been going on for a long while. The teacher wondered if her coming over to Georgia when she whined, helped support her behavior. To find out, the teacher told Georgia to raise her hand and ask for help if she was having trouble. Otherwise, Georgia would not get any help. This system worked fine. Georgia does not whine, and the whole classroom atmosphere is much better.

    Say whether the above is an example of abulia, the constructional approach or the tau effect. If it is one of these three concepts, justify your answer. If it is not one of these, rewrite the passage to make it illustrate the concept to which it is closest. Then justify the changes that you made.

12. List the factors that affect habit strength.

13. Define the "tau effect" in your own words.

14. Kim grew up in a rural setting where she would say hi to everyone she met, whether she knew them or not. People would always smile and say hello back. When she went away to college, at Hugh State University, however, she found that when she greeted people that she didn't know, not only didn't they return her greeting, but they also avoided eye contact with her. On some occasions they acted as if she was trying to pick them up. She rarely says hi to strangers now.

    Say whether the above is an example of abulia, the constructional approach, or the tau effect. If it is one of these three concepts, justify your answer. If it is not one of these, rewrite the passage to make it illustrate the concept to which it is closest. Then justify the changes that you made.

15. Give an original example of the constructional approach.

16. Say which of the following are examples of abulia:

    a. Graelle was an elephant at the Metropolitan Zoo. Her trainer decided to teach her to lift her trunk, grab a hammer, and hit a lever. At first, the trainer would give her a peanut every sixth time that she hit the lever. Graelle spent a great deal of time every day trying to get the peanuts. One day, the trainer decided to give her food everytime that Graelle hit the lever. She ate the first 100 peanuts, hitting the lever very hard every time. Then she began to slow down, and finally gave up.

    b. Andrea is reading a novel; July's Mixed Blessing, by H.W. Chartier. It is an adventure story about a 14 year old youth, Brent, who survives a canoeing accident in the wilderness of Northern Maine.

His partner was not so lucky. The book describes many of his trials and tribulations as he tries to make his way back to civilization. At first, these new and exciting perils occur in nearly every chapter, and Andrea can't put the book down. Then, in chapter 6, Chartier starts expounding upon the wildlife that Brent encounters. These descriptions continue for 5 more chapters. Andrea never gets to chapter 11.

c. Sharon is a transfer student in electrical engineering. At her previous school she hated doing research projects, mostly because it was so difficult to obtain all the books that she needed. But she kept at the projects, completing 8 papers in her first year. When she transferred to the State University, she was amazed because everytime she looked for a book, it was there. However, during her first semester, she completed only two of the four papers that she was supposed to do.

d. John bagged groceries at the local supermarket. He always remembered to put cans on the bottom and eggs on the top. He even put soap items separate from meat items. Customers often remarked about what a good worker he was. He was never late and was always willing to work over time. He even worked Saturday nights and Sunday when asked. Within a month, John was promoted to work the dairy case. Fewer customers got a chance to comment on what a good job he did. His new manager was also not so liberal with the compliments. John was still never late and did a great job stocking the dairy case. Eventually he became assistant manager of the store.

e. When Karen and Bob first started going out together they rarely went to the movies. They found that when they did, the shows were almost always a disappointment. Recently, however, they went to a couple of movies and they found them both entertaining. Now they go to the movies at least once a week.

17. Sony is learning the names of different animals. His older sister is looking through a book with him and having him name the animals on each page. She shows him a picture of a cow and he says, "cow." She shows him a picture of a cow a moment later and he says, "That's the picture you just showed me, and it's a cow." The next day they do the same thing. When she shows him a picture of a cow he says, "Gee, I think it's a walrus."

Say whether the above is an example of abulia, the constructional approach, or the tau effect. If it is one of these three concepts, justify your answer. If it is not one of these three concepts, rewrite the passage to make it illustrate the concept to which it is closest. Then justify the changes that you made.

18. Give an original example that illustrates the two variables that interact with habit strength to make a response occur.

19. Ralph always beat up his sister,Annie. He constantly hit her and bruised her somehow. Every time he did this, his mother came rushing over and told him to stop. Otherwise she was always off somewhere else in the house, either cleaning or talking with a neighbor. She became very worried when one day Ralph broke Annie's arm. She decided to seek some help at the local Family Counseling Center. She told a counselor about the situation. The counselor suggested that she may be reinforcing hitting his sister by running up and attending to him. The counselor suggested two ways to deal with the situation. First, she should make an effort to attend to the children when they are playing cooperatively. Second, if Ralph does hit Annie, his mother should put him in his room for 5 minutes, and should not say anything to him while she is bringing him to his room or while he is in his room. Since that time, Ralph has hit Annie only twice, and has not hit her for the last five weeks.

Say whether the above is an example of abulia, the constructional approach, or the tau effect. If it is one of these three concepts, justify your answer. If it is not one of these three concepts, rewrite the passage to make it illustrate the concept to which it is closest. Then justify the changes that you made.

APPENDIX F

A Comment and Scoring Sheet

DATA SHEET

Researcher ____________________ Subject ____________________

Date ____________________ Program Type ____________________

Tape ____________________

Concept ____________________

Task #	Duration	Correct	Incorrect	R..R

APPENDIX G

Answer Keys

Constructional Approach

Answer Key

Definitions - Must have each of the following features. Each feature is worth 2 points.

1. observation or interview to determine problem behavior
2. observation or interview to determine critical reinforcer
3. alternative behavior taught or suggested
4. alternative behavior obtains the same critical reinforcer
5. alternative behavior is emitted in the same environment as problem behavior

Example Request - Must conform to the features specified for definitions. Each exemplifies feature is worth 2 points.

Example Identification - See individual transfer test keys.

Combinations - Must be evaluated like a Bloom's taxonomy evaluation task. Therefore, either a yes response or a no response is correct, but the appropriate justification must follow.

If yes, then:

2 points for each feature like in the definitions.

if no, then:

1. 2 points for no
2. 4 points for justification
3. 4 points for change

Tau Effect

Answer Key

Definitions - Must have each of the following features. Each feature is worth 2 points.

1. three successive presentations of the same object or event
2. statement of identity
3. short period of time
4. statement of difference
5. long period of time

Features 2 and 3 must be related; Features 4 and 5 must be related.

Example Request - Must conform to the features specified for definitions. Each exemplified feature is worth 2 points.

Example Identification - see individual transfer test keys.

Combinations - Must be evaluated as one would evaluate an evaluation level task from Bloom's taxonomy. Therefore, either a yes or a no response is correct, but the appropriate justification must follow.

If S says yes, then:

2 points for each feature, like in the definitions, except feature 1 can be implied

If S says no, then:

1. 2 points for no
2. 4 points for justification
3. 4 points for change

ABULIA

Answer Key

Definitions - Must have each of the following features. Each feature is worth 2 points.

1. Initial high rate of response
2. Initial high proportion of reinforcers to responses
3. Then the proportion of reinforcers decreases
4. Then the frequency of the response decreases
5. Feature 3 is abrupt

Example Request - Must conform to the features specified for the definitions. Each exemplified feature is worth 2 points.

Example Identification - See individual transfer test keys

Combinations - Must be evaluated like a Bloom's taxonomy evaluation level task. Therefore, either a yes response or a no response is correct, but the appropriate justification must follow.

If yes, then:

2 points for each feature like in definitions

If no, then:

1. 2 points for no
2. 4 points for justification
3. 4 points for change

APPENDIX H

Study Behavior Questionnaire for Experiment 2

Study Behavior Questionnaire

______________________ S# ______________
(name)

____
(date)

After Study, ask the subject the following questions:

1. How many times did you reread the material? ________.

   Did you do any:

2. Outlining? ________.
3. Underlining? ________.
4. Brackets? ________.
5. Marking any other part of the material? ________.
6. Did you write down key terms, names, definitions, important points? ________

   Did you design any:

7. Fill-in questions? ________
8. Multiple-choice questions? ________.
9. Short or long essay questions? ________.
10. Did you use any other study procedures?
11. ________________________
12. ________________________
13. ________________________
14. Did you orally quiz yourself? ________
15. Did you silently quiz yourself? ________
16. Did you write answers to any written study aids you made? ________
17. Did you quiz yourself without referring to your notes, study aides, or the prose passage? ________.

After Study:

1. Please indicate the level of disturbance in this carrel during the study episode by rating any visual, auditory, or other distractions?

1	2	3	4	5
Low		Medium		High

2. Please estimate the time of the entire study episode to the nearest minute ______________.

3. Please indicate your accuracy in filling out this questionnaire:

1	2	3	4	5
Low		Medium		High

APPENDIX I

Pretest for Experiment 2

PRETEST

Read each question carefully. Answer each question as completely as you can. If you don't know the answer to a question, write DK (for Don't Know).

1. Define "constructional approach" in your own words.

2. Give an original example of "abulia."

3. Say which of the following are examples of "tau effect":

   a. Todd saw a picture of Cheryl Tiegs modeling a bathing suit in Sports Illustrated, one day in 1973. When he saw her on Cover Girl makeup ads two years later, he told his girlfriend that he had predicted that she would make it big some day. Later that day he saw her on a talk show, and commented on how lucky he was to be able to see her twice in one day.

   b. Cindy likes bikini bathing suits on men. While at the beach one afternoon last summer, she saw a guy with a blue bikini. Two weeks later, she saw another guy wearing a blue bikini, and pointed out to her friend Liz that she had seen the same suit two weeks ago and thought that it would look good on her friend Joe. When she went to the beach again the next day, Liz pointed out the same blue bikini on someone else, but Cindy said the suit she liked was light green.

   c. Mary and Rod went to see "For Pete's Sake," starring Robert Redford. The next day, they went to see "Butch Cassidy and the Sundance Kid." Rod pointed out that Robert Redford was in both movies. Three months later, they went to see "All the President's Men," which also starred Robert Redford. Mary asked, "Isn't that Robert Redford?" Rod replied, "Nah, that's Paul Newman."

   d. Teddy likes to buy crackerjacks. He gets the prizes and then gives them away to his friends.

Once he got two blue secret decoder rings in the same box. He went running to his mother and told her that he got the same prize twice. He gave one of the rings to his friend Joe. Three weeks later he was visiting Joe and asked him where he got the blue secret decoder ring. Joe told Ted that he had given it to him. Joe said, "I thought I had given you a red one."

4. Define environment in your own words.

5. Roberta is an excellent tennis player. She feels great after returning a difficult serve or hitting a shot down the line. She is usually able to make a number of such excellent shots in a game. Recently, she has started to spend more time hiking, and has found it difficult to arrange time for tennis. She has gone from playing four times a week to once a week.

   Say whether the above is an example of abulia, constructional approach, or tau effect. If it is one of these three concepts, justify your answer. If it is not one of these three concepts, rewrite the passage to make it illustrate the concept to which it is closest. Then justify the changes that you made.

6. Give an original example of the "tau effect."

7. Explain the difference between effective reaction potential and momentary effective reaction potential. Be complete.

8. In each of the examples below there is a frequency of an event mentioned. If you think the event is occurring at a high frequency write, <u>high</u>. If you think the event is occurring at a low frequency, write <u>low</u>. If there is not enough information, write <u>neither</u>. Please rely solely on the information provided in the example. Say why you labelled each the way you did.

a. Joan plays Rugby 6 days a week.
b. Tod took piano lessons 2 days a week for 5 years. Now he takes piano lessons once a week. (once a week is the frequency of interest)
c. Beth can type 120 words a minute.
d. Susan reads 20 pages an hour.
e. Mark used to cook dinner 6 nights a month. Now he cooks 12 nights a month. (12 nights a month is the frequency of interest).

9. Give an original example of a time interval that is short.

10. Define abulia in your own words.

11. Say which of the following are examples of the "constructional approach":

a. Bob was a top student in highschool, although he rarely studied. Things are a little different now that he has come to college. Every night after supper, the guys talk him into playing cards. Often the card games lead to philosophical discussions, which he admittedly loves. As a result, he spends every night playing cards for about three hours, and then expounding upon philosophical issues for another three hours. This leaves little time for homework. After getting a 1.3 cum semester, Bob decides to see his advisor for suggestions. Together they decide that since he likes philosophizing so much, it might be best for him to arrange his schedule so that he studies right after supper, and discusses philosophy only after he has completed at least 90% of his homework. Since this arrangement, he has handed in nearly all of his homework on time, and has gotten above 90% on all of his exams and quizzes.

b. Arthur disrupted his third grade class by swearing a great deal. His teacher decided to have a talk with him and the principal. They concluded that the shocked face of the teacher and the giggles of the class were encouraging Arthur to swear. The teacher talked to the class and asked them not to energy policy. Kim told Carl how she was glad to hear that Carter has expanded his energy policy.

Say whether the above is an example of abulia, constructional approach, or the tau effect. If it is one of these three concepts, justify your answer. If it is not one of these three concepts, rewrite the passage to make it illustrate the concept to which it is closest. Then justify the changes that you made.

15. Say which of the following are examples of reinforcement. Say why you chose those that you did.

a. Henrico doesn't like to go to parties. Therefore whenever his girlfriend wants to go to a party he puts up a big stink, yelling and screaming about what a bore parties are. His girl friend then plays up to him, looks at him with her beautiful brown eyes and convinces him to go to the party. However, the next party that comes along, he puts on the same act. (look at Henrico's complaining behavior).

b. Molly wound her watch back and forth every morning for fourteen years. And every day for fourteen years her watch kept good time. Therefore, she continues to wind her watch back and forth.

c. Joel took photographs of beach scenes one summer. When he returned to the city in the fall he sold every single picture. In fact, someone even published a book of his pictures. He has never taken a photograph since.

d. Alphonse wrote a personal add in the <u>Valley Avocado</u>. The add said that he was looking for a mature woman who might be interested in sharing an apartment with a sensitive, but shy guy. Al got many responses, but none of the woman were exactly his type of roommate. Therefore, he put another add in the paper. Again he got many responses, but no results. Well, Al has continued to put adds in the paper at least once a month.

16. Georgia, a three year old attending a local preschool, had a history of whining when things didn't go her way. For example, one day when she was building a house out of blocks, she accidently knocked it over. She immediately started whining, and her teacher came right over. This sort of behavior had been going on for a long while. The teacher wondered if her coming over to Georgia when she whined, helped support her behavior. To find out, the teacher told Georgia to

raise her hand and ask for help if she was having trouble. Otherwise, Georgia would not get any help. This system worked fine: Georgia does not whine, and the whole classroom atmosphere is much better.

Say whether the above is an example of abulia, the constructional approach, or the tau effect. If it is one of these three concepts, justify your answer. If it is not one of these, rewrite the passage to make it illustrate the concept to which it is closest. Then justify the changes that you made.

17. Give an original example of a high frequency event.

18. List the factors that affect habit strength.

19. Define the "tau effect" in your own words.

20. Say which of the following are examples of environments: Say why.

a. eating
b. a guy
c. typing
d. a lobby
e. mountains
f. Skinner box
g. mowing
h. parties

21. Give an original example of the constructional approach.

22. Say which of the following are examples of abulia:

giggle anymore. She informed them that if they giggled when Arthur swore, they would not be able to have recess. She also tried not to look shocked anymore. This program effectively reduced Arthur's swearing; he has not sworn in class for over eight weeks.

c. Greg was mentally retarded and participated in a workshop to learn to assemble transistor radio parts. He spent much of his time distracting the other members of the workshop. They often enjoyed the distraction, and kidded around with him. This caused the supervisor to believe that the kidding around supported Greg's distractive behavior. The supervisor decided that Greg should only be allowed to kid around with the other participants if he was on task for 15 straight minutes. This proved to be much better for all concerned, as Greg quickly learned to work diligently for 15 minutes and kid around for 5.

d. Mrs. Johnson thought that Shawn talked too much in class. She decided to put an end to this verbosity once and for all. So Mrs. Johnson had the principal call Shawn down to the office to run an errand. Then, she told the class that they were to stop attending to Shawn when she talked for more than two minutes straight, and when she talked more than three times within each class period. She asked them to make sure that they continued to acknowledge Shawn's comments up to these points.

12. Define consequences in your own words.

13. Give an original example that illustrates how stimulus generalization affects effective habit strength.

14. Kim was reading the newspaper the other day and noticed a little blurb about Carter's energy policy. The next day, she saw another article on Carter's energy policy, and told her friend Carl that there had been two separate articles intwo days about Carter's energy policy. A month later, in a STATEMENT-OF-THE-UNION address, Carter expounded further on his

a. Graelle was an elephant at the Metropolitan Zoo. Her trainer decided to teach her to lift her trunk, grab a hammer, and hit a lever. At first, the trainer would give her a peanut every sixth time that she hit the lever. Graelle spent a great deal of time every day trying to get the peanuts. One day, the trainer decided to give her food everytime that Graelle hit the lever. She ate the first 100 peanuts, hitting the lever very hard every time. Then she began to slow down, and finally gave up.

b. Andrea is reading a novel, July's Mixed Blessing, by H.W. Chartier. It is an adventure story about a 14 year old youth, Brent, who survives a canoeing accident in the wilderness of Northern Maine. His partner was not so lucky. The book describes many of his trials and tribulations as he tries to make his way back to civilization. At first, these new and exciting perils occur in nearly every chapter, and Andrea can't put the book down. Then, in chapter 6, Chartier starts expounding upon the wildlife that Brent encountered. These descriptions continue for 5 more chapters. Andrea never gets to chapter 11.

c. Sharon is a transfer student in electrical engineering. At her previous school she hated doing research projects, mostly because it was so difficult to obtain all the books that she needed. But she kept at the projects, completing 8 papers in her first year. When she transferred to the State University, she was amazed because every time she looked for a book, it was there. However, during her first semester, she completed only two of the four papers that she was supposed to do.

d. John bagged groceries at the local supermarket. He always remembered to put cans on the bottom and eggs on the top. He even put soap items separate from meat items. Customers often remarked about what a good worker he was. He was never late and was always willing to work overtime. He even worked Satruday nights and Sunday when asked. Within a month, John was promoted to work the dairy case. Fewer customers got a chance to comment on what a good job he did. His new manager was also not so liberal with the compliments. John was still never late and did a great job stocking the dairy case. Eventually he became assistant manager of the store.

23. Give an original example of reinforcement.

24. Say what the consequences are in the following examples. Say why you labelled these as consequences.

    a. John has an eating problem. Whenever he tries to stick a fork full of peas in his mouth they fall off the fork. His brothers always laugh when this happens.
    b. Reading a magazine the other day, I came across an add for Ron Rico Rum. It was so obnoxious that I took the bottle of Ron Rico Rum that I had in my closet and broke it.
    c. Martha was writing a letter home to her father. Her pen suddenly ran out of ink. So, she went and put another cartridge in it.
    d. Billie caught her arm in a crack on the side of a cliff. She tried to loosen it by pulling herself up, but to no avail. She had to hang there for four hours before help came.

25. Given an original example that illustrates the two variables that interact with habit strength to make a response occur.

26. Give an original example of a consequence.

27. What information would you need to know in order to say whether some events occur at a high or a low frequency?

28. What information would you need in order to say whether a time interval is long or short?

29. Define reinforcement in your own words.

30. In each of the examples below there is a time interval between some events. If you think that this time interval is long, write long. If you think that this time interval is short, write short. If you think that there is not enough information to decide, write neither. Please rely solely on the information provided in the example. Say why you answered each the way you did.

a. Tammy took a cottage on the beach for a weekend in January. In May she rented the same cottage for two nights. Finally, she rented the cottage in November for a weekend. (time between May and November is the time interval of interest)

b. Dick caught a pass in the first quarter of yesterdays game. Then he caught a second pass with one minute to go in the half. (time interval between first quarter and last minute of half is the time of interest)

c. One day while walking down the street I saw a dark pink Rambler fly by at 100 miles per hour. Six weeks later I saw the very same Rambler fly by at close to 100 miles per hour. (time between first sighting of Rambler and second sighting is time of interest)

d. Karen took a picture of Brenden with a Pronto Instant Camera. The picture took five minutes to develop completely. Then she took a picture of Brenden with an SX-70. This picture took 2 minutes to develop completely. (time interval between taking second picture and complete development is the time of interest)

31. Give an original example of an environment.